USENIX Association

# Proceedings of the Fifteenth Symposium on Usable Privacy and Security


August 12–13, 2019
Santa Clara, CA, USA

# Symposium Organizers

**General Chair**
Heather Richter Lipford, *University of North Carolina at Charlotte*

**Invited Talks Chair**
Serge Egelman, *International Computer Science Institute*

**Technical Papers Co-Chairs**
Michelle Mazurek, *University of Maryland*
Rob Reeder, *Google*

**Technical Papers Committee**
Yasemin Acar, *Leibniz Universität Hannover*
Adam Aviv, *United States Naval Academy*
Lujo Bauer, *Carnegie Mellon University*
Konstantin Beznosov, *The University of British Columbia*
Tamara Bonaci, *Northeastern University and University of Washington*
Joe Calandrino, *Federal Trade Commission*
Sonia Chiasson, *Carleton University*
Nicolas Christin, *Carnegie Mellon University*
Lynne Coventry, *Northumbria University*
Serge Egelman, *International Computer Science Institute (ICSI) and University of California, Berkeley*
Sascha Fahl, *Ruhr-Universität Bochum*
Carrie Gates, *Bank of America*
Rachel Greenstadt, *New York University*
Iulia Ion, *Google*
Apu Kapadia, *Indiana University Bloomington*
Bart Knijnenburg, *Clemson University*
Katharina Krombholz, *Helmholtz Center for Information Security (CISPA)*
Susan McGregor, *Columbia University Graduate School of Journalism*
Emilee Rader, *Michigan State University*
Scott Ruoti, *The University of Tennessee, Knoxville*
Florian Schaub, *University of Michigan*
Kent Seamons, *Brigham Young University*
Manya Sleeper, *Google*
Matthew Smith, *University of Bonn*
Jessica Staddon, *Google*
Elizabeth Stobert, *National Research Council Canada*
Jose Such, *King's College London*
Nina Taft, *Google*
Mary Theofanos, *National Institute of Standards and Technology (NIST)*
Blase Ur, *The University of Chicago*
Kami Vaniea, *The University of Edinburgh*
Emanuel von Zezschwitz, *University of Bonn and Fraunhofer FKIE*
Rick Wash, *Michigan State University*

**Lightning Talks and Demos Co-Chairs**
Yousra Javed, *Illinois State University*
Scott Ruoti, *University of Kentucky*

**Karat Award Chair**
Kent Seamons, *Brigham Young University*

**Posters Co-Chairs**
Yasemin Acar, *Leibniz University Hannover*
Heather Crawford, *Florida Institute of Technology*

**Tutorials and Workshops Co-Chairs**
Elissa Redmiles, *University of Maryland*
Daniel Zappala, *Brigham Young University*

**Publicity Co-Chairs**
Nalin Asanka Gamagedara Arachchilage, *University of New South Wales*
Joe Calandrino, *Federal Trade Commission*

**Program Committee Local Arrangements Chair**
Kami Vaniea, *University of Edinburgh*

**Email List Chair**
Lorrie Cranor, *Carnegie Mellon University*

**Accessibility Chair**
Rich Williams, *USENIX Association*

**USENIX Liaison**
Casey Henderson, *USENIX Association*

**Steering Committee**
Robert Biddle, *Carleton University*
Sonia Chiasson, *Carleton University*
Sunny Consolvo, *Google*
Lorrie Cranor, *Carnegie Mellon University*
Patrick Gage Kelley, *Google*
Jaeyeon Jung, *Samsung Electronics*
Apu Kapadia, *Indiana University Bloomington*
Michelle Mazurek, *University of Maryland*
Rob Reeder, *Google*
Mike Reiter, *University of North Carolina, Chapel Hill*
Heather Richter Lipford, *University of North Carolina at Charlotte*
Matthew Smith, *University of Bonn, Fraunhofer FKIE*
Rick Wash, *Michigan State University*

# External Reviewers

Reham Mohamed

Hala Assal

Fiona Westin

Sandra Gabriele

Ben Morrison

James Nicholson

Matthias Fassl

Tousif Ahmed

Taslima Akter

Roberto Hoyle

Heather Molyneaux

Artemij Voskobojnikov

Borke Ejaita Obada-Obieh

Masoud Mehrabi Koushki

Yue Huang

# SOUPS 2019
# Fifteenth Symposium on Usable Privacy and Security
# Message from the Chairs

Welcome to SOUPS 2019!

As SOUPS turns 15, we are happy to see that the conference continues to thrive. We are pleased this year to present a program that covers a broad range of topics within usable privacy and security. Technical paper presentations form the core of the SOUPS program, but the conference also includes workshops, tutorials, posters, lightning talks, and a keynote.

In 2016, SOUPS became an independent conference body. For the last three years, we have partnered with USENIX for hosting and administrative support, a move that has enabled continued growth for the conference. We thank all the members of the USENIX staff for their work in organizing SOUPS and supporting our community. In 2018, we were co-located with the USENIX Security Symposium for the first time, and we are continuing that co-location for 2019, 2020, and 2021. Co-locating the two conferences allows for interactions and shared ideas between SOUPS and USENIX Security attendees; we found this beneficial for both conferences last year and look forward to repeating it this year, in Boston in 2020, and in Vancouver, BC, in 2021.

SOUPS relies on a range of volunteers for all of its activities. Steering Committee members provide oversight and guidance and are elected for three year terms. Organizing Committee members help determine the conference content for a particular year, often serving two year terms to facilitate the transition of knowledge. Technical Papers Committee members are chosen by the Technical Papers co-Chairs each year. SOUPS is a product of the hard work by all the SOUPS Organizers, the SOUPS Steering Committee, the Technical Papers Committee, the Tutorial and Workshop organizers, the Posters jury, and the USENIX staff. We thank each and every one of you for your contributions to SOUPS 2019.

Heather is serving her first year as General Chair of SOUPS and Chair of the Steering Committee. Next year, Heather will serve again as General Chair and will work with Sonia Chiasson, who will serve as Vice Chair in 2020, and General Chair in 2021 and 2022 after that. If you are interested in helping with SOUPS 2020 in any way, please contact Heather or Sonia. We are also pleased to announce that Joe Calandrino has agreed to serve as Technical Papers Co-Chair for 2020 and 2021.

We thank our sponsors, Facebook, Google, and Mozilla. SOUPS would not be possible without their generous support. Please visit our website to view the recipients of the SOUPS 2019 awards—Distinguished Paper, IAPP SOUPS Privacy Award, Distinguished Poster, and the John Karat Usable Privacy and Security Student Research Award. Congratulations to all of the recipients for their outstanding work.

Heather Richter Lipford, *University of North Carolina at Charlotte*
General Chair

Michelle Mazurek, *University of Maryland*
Technical Papers Co-Chair

Rob Reeder, *Google*
Technical Papers Co-Chair

# SOUPS 2019
# Fifteenth Symposium on Usable Privacy and Security

## August 12–13, 2019
## Santa Clara, CA, USA

### Populations and Scales

**Cooperative Privacy and Security: Learning from People with Visual Impairments and Their Allies . . . . . 1**
Jordan Hayes, Smirity Kaushik, Charlotte Emily Price, and Yang Wang, *Syracuse University*

**Privacy and Security Threat Models and Mitigation Strategies of Older Adults . . . . . 21**
Alisa Frik, *International Computer Science Institute* (*ICSI*) *and University of California, Berkeley;* Leysan Nurgalieva, *University of Trento;* Julia Bernd, *International Computer Science Institute* (*ICSI*)*;* Joyce Lee, *University of California, Berkeley;* Florian Schaub, *University of Michigan;* Serge Egelman, *International Computer Science Institute* (*ICSI*) *and University of California, Berkeley*

**Evaluating Users' Perceptions about a System's Privacy: Differentiating Social and Institutional Aspects. . . . . 41**
Oshrat Ayalon and Eran Toch, *Tel Aviv University*

**A Self-Report Measure of End-User Security Attitudes (SA-6) . . . . . 61**
Cori Faklaris, Laura Dabbish, and Jason I. Hong, *Carnegie Mellon University*

### Security Behaviors and Experiences

**The Effect of Entertainment Media on Mental Models of Computer Security . . . . . 79**
Kelsey R. Fulton, Rebecca Gelles, Alexandra McKay, Richard Roberts, Yasmin Abdi, and Michelle L. Mazurek, *University of Maryland*

**A Typology of Perceived Triggers for End-User Security and Privacy Behaviors . . . . . 97**
Sauvik Das, *Georgia Institute of Technology;* Laura A. Dabbish and Jason I. Hong, *Carnegie Mellon University*

**Replication: No One Can Hack My Mind Revisiting a Study on Expert and Non-Expert Security Practices and Advice . . . . . 117**
Karoline Busse and Julia Schäfer, *University of Bonn;* Matthew Smith, *University of Bonn/Fraunhofer FKIE*

***"Something isn't secure, but I'm not sure how that translates into a problem":* Promoting autonomy by designing for understanding in Signal. . . . . 137**
Justin Wu, Cyrus Gattrell, Devon Howard, and Jake Tyler, *Brigham Young University;* Elham Vaziripour, *Utah Valley University;* Kent Seamons and Daniel Zappala, *Brigham Young University*

**"I was told to buy a software or lose my computer. I ignored it": A study of ransomware . . . . . 155**
Camelia Simoiu, *Stanford University;* Christopher Gates, *Symantec;* Joseph Bonneau, *New York University;* Sharad Goel, *Stanford University*

### New Paradigms

**Enhancing Privacy through an Interactive On-demand Incremental Information Disclosure Interface: Applying Privacy-by-Design to Record Linkage . . . . . 175**
Hye-Chung Kum, *Population Informatics Lab, Texas A&M University;* Eric D. Ragan, *INDIE Lab, University of Florida;* Gurudev Ilangovan, Mahin Ramezani, Qinbo Li, and Cason Schmit, *Population Informatics Lab, Texas A&M University*

**From Usability to Secure Computing and Back Again . . . . . 191**
Lucy Qin, Andrei Lapets, Frederick Jansen, Peter Flockhart, Kinan Dak Albab, and Ira Globus-Harris, *Boston University;* Shannon Roberts, *University of Massachusetts Amherst;* Mayank Varia, *Boston University*

**Certified Phishing: Taking a Look at Public Key Certificates of Phishing Websites . . . . . 211**
Vincent Drury and Ulrike Meyer, *Department of Computer Science, RWTH Aachen University*

## Developers and Sysadmins

**"We Can't Live Without Them!" App Developers' Adoption of Ad Networks and Their Considerations of Consumer Risks . . . . . 225**
Abraham H. Mhaidli, Yixin Zou, and Florian Schaub, *University of Michigan School of Information*

**Usability Smells: An Analysis of Developers' Struggle With Crypto Libraries . . . . . 245**
Nikhil Patnaik, Joseph Hallett, and Awais Rashid, *University of Bristol*

**System Administrators Prefer Command Line Interfaces, Don't They? An Exploratory Study of Firewall Interfaces . . . . . 259**
Artem Voronkov, Leonardo A. Martucci, and Stefan Lindskog, *Karlstad University*

**Keepers of the Machines: Examining How System Administrators Manage Software Updates . . . . . 273**
Frank Li, *University of California, Berkeley;* Lisa Rogers, *University of Maryland;* Arunesh Mathur, *Princeton University;* Nathan Malkin, *University of California, Berkeley;* Marshini Chetty, *Princeton University*

## Authentication

**Communicating Device Confidence Level and Upcoming Re-Authentications in Continuous Authentication Systems on Mobile Devices. . . . . 289**
Lukas Mecke, *University of Applied Sciences Munich, Munich, Germany and LMU Munich, Munich, Germany;* Sarah Delgado Rodriguez and Daniel Buschek, *LMU Munich, Munich, Germany;* Sarah Prange, *University of Applied Sciences Munich, Munich, Germany and Bundeswehr University Munich, Munich, Germany and LMU Munich, Munich, Germany;* Florian Alt, *Bundeswehr University Munich, Munich, Germany*

**Exploring Intentional Behaviour Modifications for Password Typing on Mobile Touchscreen Devices. . . . . 303**
Lukas Mecke, *University of Applied Sciences Munich, Munich, Germany and LMU Munich, Munich, Germany;* Daniel Buschek and Mathias Kiermeier, *LMU Munich, Munich, Germany;* Sarah Prange, *University of Applied Sciences Munich, Munich, Germany and Bundeswehr University Munich, Munich, Germany and LMU Munich, Munich, Germany;* Florian Alt, *Bundeswehr University Munich, Munich, Germany*

**Why people (don't) use password managers effectively . . . . . 319**
Sarah Pearman, Shikun Aerin Zhang, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor, *Carnegie Mellon University*

**Of Two Minds about Two-Factor: Understanding Everyday FIDO U2F Usability through Device Comparison and Experience Sampling. . . . . 339**
Stéphane Ciolino, *OneSpan Innovation Centre & University College London;* Simon Parkin, *University College London;* Paul Dunphy, *OneSpan Innovation Centre*

**A Usability Study of Five Two-Factor Authentication Methods. . . . . 357**
Ken Reese, Trevor Smith, Jonathan Dutson, Jonathan Armknecht, Jacob Cameron, and Kent Seamons, *Brigham Young University*

## Personal Privacy

**Personal Information Leakage by Abusing the GDPR "Right of Access" . . . . . 371**
Mariano Di Martino and Pieter Robyns, *Hasselt University/tUL, Expertise Centre For Digital Media;* Winnie Weyts, *Hasselt University - Law Faculty;* Peter Quax, *Hasselt University/tUL, Expertise Centre For Digital Media, Flanders Make;* Wim Lamotte, *Hasselt University/tUL, Expertise Centre For Digital Media;* Ken Andries, *Hasselt University - Law Faculty, Attorney at the Brussels Bar*

**An Empirical Analysis of Data Deletion and Opt-Out Choices on 150 Websites. . . . . 387**
Hana Habib, *Carnegie Mellon University;* Yixin Zou, *University of Michigan;* Aditi Jannu, Neha Sridhar, Chelse Swoopes, Alessandro Acquisti, Lorrie Faith Cranor, and Norman Sadeh, *Carnegie Mellon University;* Florian Schaub, *University of Michigan*

**The Fog of Warnings: How Non-essential Notifications Blur with Security Warnings. . . . . 407**
Anthony Vance, *Temple University;* David Eargle, *University of Colorado Boulder;* Jeffrey L. Jenkins, C. Brock Kirwan, and Bonnie Brinton Anderson, *Brigham Young University*

(continued on next page)

## Wearables and Smart Homes

**"There is nothing that I need to keep secret": Sharing Practices and Concerns of Wearable Fitness Data . . . . . 421**
Abdulmajeed Alqhatani and Heather Richter Lipford, *University of North Carolina at Charlotte*

**"I don't own the data": End User Perceptions of Smart Home Device Data Practices and Risks . . . . . 435**
Madiha Tabassum, *University of North Carolina at Charlotte;* Tomasz Kosinski, *Chalmers University of Technology;* Heather Lipford, *University of North Carolina at Charlotte*

**More than Smart Speakers: Security and Privacy Perceptions of Smart Home Personal Assistants . . . . . 451**
Noura Abdi, *King's College London;* Kopo M. Ramokapane, *University of Bristol;* Jose M. Such, *King's College London*

# Cooperative Privacy and Security: Learning from People with Visual Impairments and Their Allies

Jordan Hayes
*Syracuse University*

Smirity Kaushik
*Syracuse University*

Charlotte Emily Price
*Syracuse University*

Yang Wang
*Syracuse University*

## Abstract

To better inform privacy/security designs for people with disabilities, we "shadowed" people with visual impairments and their allies (e.g., friends, family members, and professional helpers) for two days followed by an exit interview. Our study results provide rich and nuanced accounts of how people with visual impairments enact their privacy/security in daily life, influenced by both their interactions with their allies and multiple (marginalized) dimensions of their identities such as different disabilities. We also found that people with visual impairments often work closely with their allies to protect their privacy and security in a cooperative manner. However, they were also thoughtful about who they would ask for help in part due to privacy reasons, even if they are trustworthy family members. We discuss ideas for future research and design, particularly a need for designing mechanisms or tools that facilitate cooperative privacy management (e.g., between people with visual impairments and their allies).

## 1 Introduction

The majority of end-user privacy/security mechanisms rely on visual cues, such as checking the lock icon for secure web connections (HTTPS), and scanning the environment for physical security threats. These approaches are challenging for people with visual impairments, which include people on a spectrum ranging from low vision to complete vision loss, and in some cases co-existing with other disabilities. We also agree that "disabilities need not to be fixed but are assets in their own right" [45]. Historically, disability is defined by a "lack of ability, knowledge, etc" and often technologies are seen as a means to fix this so called "lack of." Instead, we are challenging this notion by looking at the experiences of disability as socially constructed and valuable to improving technologies and systems rather than the opposite.

Our long-term research goal is to design effective privacy/security mechanisms to better support people with disabilities. To help inform future design, we conducted an observational study involving adults with visual impairments and their allies (e.g., friends, family members, professional helpers) to answer two main research questions:

- RQ1. What are the everyday privacy/security challenges and practices of people with visual impairments?
- RQ2. How do people with visual impairments interact with their allies? What are the privacy or security implications of such interactions?

We use the term *ally* to explore the complexities of social relationships between people with disabilities and those who respect and often interact with them. We use ally rather than caregiver because the latter implies a one-sided relationship whereas the former implies "equality, mutual trust, and shared decision-making" [20]. In our research, we sought to bring a marginalized group to the center [17] and therefore individuals with visual impairments were the primary focus of our study. We also explored their relationships and interactions with allies, many of whom also participated in our study.

Compared with prior work on privacy/security practices of people with visual impairments (e.g., [3,4]), the novelty of our research is twofold. First, from a methodological perspective, we employed an observational technique (i.e., "*shadowing*" participants [34, 48]), complemented with semi-structured interviews to understand participants' lived everyday experiences. Analytically, we paid special attention to how our participants' everyday privacy/security experiences are shaped by their interactions with allies and by different (marginalized) dimensions of their identities such as disability and gender identity ( [35]). Second, our study provides novel findings


*USENIX Symposium on Usable Privacy and Security (SOUPS)* 2019.
August 11–13, 2019, Santa Clara, CA, USA.

regarding participants' views of their own visual impairments and their conceptualizations of privacy. Both may influence their everyday privacy and security practices (e.g., disability disclosure and willingness to ask for help). We also uncover understudied privacy/security challenges participants faced (e.g., managing social relationships, sharing information with their allies under organizational policies). Our results also suggest that privacy was a key factor in considering who they would ask for help, including trusted family members.

Our findings lead to two key takeaway messages for future privacy research and design. First, *privacy management can have an inherently cooperative dimension*. As our study highlights, people with visual impairments often work closely with their allies to protect their privacy and security. Second, *privacy tools designed with underserved users in mind should pay attention to multiple marginalized identities these users might have.* As we saw in our study, some participants have other disabilities or marginalized identities. Their lived experiences are not only influenced by a single identity (e.g., visual impairments) but by the intersecting effect of all marginalized identities. Designing mechanisms to support this type of cooperative privacy management while considering the various marginalized identities of underserved users is a worthwhile direction for future research.

## 2 Related Work

In this section, we review prior work on privacy/security challenges for people with visual impairments and the roles that allies play in the lives of marginalized groups.

### 2.1 People with Visual Impairments

A few studies have focused on privacy/security issues for people with visual impairments. Many privacy/security threats arise from the use of accessible technology, as these devices inadvertently generate new avenues for passersby to learn personal information. People with visual impairments have concerns about aural and visual eavesdropping in public when using screen readers and screen magnifiers, respectively [7, 19, 30, 37]. Prior work also suggests that this user group may not notice privacy/security risks in their environment or inherent in the technology they use [14]. The use of accessible technologies can also draw unwanted attention and potential exploitation [44]. To mitigate some of these issues, people with visual impairments use privacy features (e.g., iOS Screen Curtain) and wear headphones to mitigate problems with screen readers [53]. Ahmed et al. identified privacy/security concerns or challenges people with visual impairments face such as difficulties verifying the security of banking or shopping websites, maintaining privacy on social media, asking strangers for help [3] and physical safety/security challenges in public spaces and at home [4]. Our study sheds new light on how people with visual impairments think about privacy as well as how their privacy/security experiences are shaped by their (marginalized) identity dimensions (e.g., multiple disabilities) and interactions with allies.

### 2.2 The Roles of Allies

People with visual impairments, particularly those who are blind, often have allies who assist them in various aspects of their lives. Previous research suggests that various challenges for people with visual impairments [32] are impacted by allies. For instance, Paisios et al. note the ways in which the visual nature of banking has left people with visual impairments reliant on the kindness of strangers, which can ultimately be a risk and fail to allow for independence from these individuals [39]. Kröger explores the intersection of ally research and disability theories and suggests treating both people with visual impairments and their allies as participants, exploring how their relationship can contribute to structural issues including autonomy [32]. Prior research has also suggested allies' concerns about maintaining the privacy of those they interact with. For instance, in an interview study of older adults and their allies, the latter were found to heavily rely on routines (e.g., older adults sending daily email check-ins) to manage the tension between their need for awareness and the older adults' need for privacy [11]. Systems that enable routines aligned with the allies' goals could, therefore, receive wider acceptance [11]. Prior research also suggests people with visual impairments are keenly aware of being perceived as a burden, and may wish not to depend on friends, family members, or other people [15]. While these studies provide valuable insight into how some underserved groups and allies collaborate, the privacy-related practices and implications of allies interacting with people with visual impairments remain understudied. Our results include observations of adults with visual impairments and their allies, shedding light on the privacy/security implications of their interactions.

## 3 Study Methodology

Unlike previous research, which was mostly based on interviews, this study employed two rounds of passive observations during a weekday and a weekend day, focusing on how our participants go about their everyday lives. We believe that our observational approach is appropriate because this method strikes a good balance between the acceptability/manageability for our participants to be observed as well as grasping their diverse and multi-faceted lived experiences embedded in rich social settings. We also conducted semi-structured interviews with our participants to complement the observations. The study has been approved by our IRB.

### 3.1 Participant Recruitment

From August 2017 to April 2018, we recruited adult participants in a metropolitan area in the Northeastern part of the US via email, phone, flyers, newsletters affiliated with local disability organizations, and attending monthly meetings with a local group of people with visual impairments. We also encouraged participants to recruit their friends and/or family members via snowball sampling [10]. Due to the nature of our study (e.g., extensive observation), we found recruiting participants to be a significant challenge. We recognize our small sample size as a key limitation, which prevents us from generalizing our results. However, similar to qualitative studies in general, our study provides rich accounts of people's lived experiences, which large-scale quantitative studies hardly offer. We were interested in the "richness" of their experiences rather than the commonality of these experiences. We created a pro-rate scheme to compensate our participants up to $60: initial interview ($5), first observation ($20), second observation ($20), exit interview ($5), and full completion of study tasks ($10). Both our participants with visual impairments and ally participants read and signed consent forms before the study. All participants completed the entire study.

### 3.2 Build Trust with Disability Communities

For more than a year before this study, members of our team were invited to and have been regularly attending monthly gatherings of a local group of people with visual impairments. Our research team has also volunteered in many social events hosted by this group (e.g., BBQ parties). In addition, we were invited to subscribe to their email mailing list. With their permission, we used this mailing list to distribute our recruitment materials to the group. All interactions with this group allowed us to get to know each other and build trust with them. Three of our eight participants came from this group. We also attended meetings involving a group of local military veterans with visual impairments and presented this study to them. Two participants came from this veteran group.

### 3.3 Other Ethical Considerations

Following suggestions from prior work [27], we told our participants that they could pause our observations and/or quit the study at any time. Upon participants' explicit permission, we audio recorded every session. We also verbally notified each participant when we started each audio recording. We allowed the recording to continue throughout the entire session, including the interview and observation portions of each study session. The only exceptions included travel between observation sites (e.g., from work to home) or our observations of participants conducting an outside activity (e.g, walking in a park). To ensure correct understanding of participants, we reviewed the main points we learned from our participants to clarify any misunderstandings we might have. We also sent a draft of our paper to our participants for feedback.

### 3.4 Study Protocol

We set out to study both adults with visual impairments and their allies. We told our prospective participants that we are interested in learning their experiences in daily life in order to understand their privacy/security needs and to inform technology designs that can better support people with visual impairments. This study has five main components: an initial interview, a weekday session of observation, a weekend session of observation, a set of pre-defined tasks in one observation session, and an exit interview. We include the study script in the Appendix. In all cases, two researchers conducted the study: one led the session and another took detailed notes.

The study started with an initial interview, in which participants were asked to describe themselves including their demographics (e.g., age bracket, gender identity, occupation), disability identity if they had any, what their daily life looks like (e.g., normal schedules, activities), and their general experiences with computing technologies and the Internet (e.g., computers, mobile devices, accessible technologies). We also asked allies about their relationships with our individuals with visual impairments and various tasks with which they often assist. This initial interview provided us relevant background information about our participants.

We then conducted observations of participants for a few hours per session (ranging from 2 to 8 hours, ending upon participants' requests): one during the week and another on the weekend, respectively. This ensured that we could learn about their lives and practices at home and work (if applicable). The observations were done using a *shadowing* method, which means that researchers follow and observe participants while taking note of what participants were saying or doing. Shadowing has been demonstrated as an effective way of gathering rich and in-depth qualitative data in the HCI community [34,48]. Since our goal was to gain a deep understanding of participants' privacy/security practices in everyday life, shadowing allows us to embed ourselves in the cultural and social settings of our participants' lives. We attempted to make our shadowing as unobtrusive as possible and informed our participants to behave as they normally would. When possible, we shadowed individuals with visual impairments and their allies simultaneously. However, at other times we shadowed participants with visual impairments without their allies due to scheduling/availability issues associated with them.

Throughout each observation, we took detailed notes using a template we designed. This template included fields for the start/end times and location of each activity we had observed, presence of other people, use of commodity and accessibility technologies; involvement of others (e.g., allies) in the activity; any privacy or security challenges encountered; any information being exchanged between individuals with

visual impairments and their allies and any hesitation each participant expressed in seeking help or relaying information.

We observed our participants conducting various activities such as grocery shopping, walking in a park, depositing money, checking mails, giving lectures, working on assignments, playing online games, using social media, receiving mobility training, doing laundry and using home appliances.

At the end of the second session, we asked our participants with visual impairments to perform or demonstrate a short list of pre-defined tasks. These tasks prompted participants to demonstrate how they use their technologies everyday, focusing on computers and/or mobile phones. A few examples included using their email and social media and demonstrating use of accessible technologies. We asked our ally participants to describe in detail how they assist individuals with visual impairments with the activities they reported in the initial interview. We wrote memos reflecting what we learned from each session and met regularly to discuss our notes.

Lastly, during exit interviews, we inquired further about the experiences of our participants with visual impairments in asking for assistance. We probed allies specifically about potential access to more personal information than needed, feelings of regret in providing help, and times when they became responsible for the privacy of those they work with. We then asked all participants a short set of questions about their privacy attitudes, concerns, and solutions. We also asked them for feedback about the study.

### 3.5 Analysis

Field notes and interviews were the primary sources of data for this study. We used memoing [25] to reflect our observations and conducted a thematic analysis [13]. After collecting and reviewing all audio recordings and participant interactions, we held weekly meetings to discuss our notes. Four researchers independently reviewed our collected data multiple times to gain a general understanding of each participants' experiences. Next, each researcher completed a round of open coding for the same two participants, deriving codes directly from the data rather than applying an established theory. We then collectively discussed the individual coding schemes and converged on a shared codebook to code the rest of our data. Next, we had another round of group discussion to walk through every participant's data and coding. We then explored higher-order connections between codes using affinity diagrams [16]. We identified notable themes, such as participants' views of their disabilities, their definitions of privacy, their awareness of/response to security and privacy threats, and relationships and interactions with allies in their everyday lives. We then had another set of group meetings to discuss and interpret the examples from our study with an eye on underlying factors such as agency and trust.

### 3.6 Participant Background

We had a total of eight participants, including three blind participants (P1, P2, P5) and two participants with low vision (P3, P4) as well as three allies (A1-P1, A2-P2, A3-P5, allies of P1, P2, and P5, respectively). We asked P3 if he was comfortable with us asking his wife (ally) to be part of the study, but he refused because he wanted to be as independent as possible. Since P4's ally (her daughter) is not an adult, we cannot have her as a participant due to our protocol. We conducted a total of 13 observation sessions, where during six sessions both participants with visual impairments and their allies were present. Table 1 summarizes the demographics of our participants. Table 2 shows the time, location, and whether allies were present at each study session.

## 4 Privacy/Security Perceptions and Practices

In this section, we first focus on how our participants with visual impairments viewed their disabilities, and how they thought about what privacy and security meant to them. These perceptions influenced their behaviors. We will then present how they dealt with their privacy and security both online and offline in their everyday lives.

### 4.1 Self-Perceptions of Their Disabilities

To understand the everyday experiences of our participants with visual impairments, we observed how they viewed their disabilities. We checked with our participants to verify these descriptions. Self-perceptions are important because they can influence our participants' behaviors, which may have significant privacy and security implications such as hiding or concealing their disability identities because of perceived stigma; this is known as visibility of disability identity [21].

P1 is a student in a US university and is originally from Tanzania. He was born blind and lives alone in an apartment close to campus[1] He has a part-time job in which he helps with issues associated with accessible technologies. He often hangs out with a friend from work (our participant, A1-P1, an African American office manager). He sometimes asks A1-P1 for help, e.g.,grocery shopping. P1 was open and accepting of his disability and has a preference for independence. For instance, instead of asking A1-P1 for help, P1 has recently started using Uber to go places by himself.

P2 is a Caucasian Reiki master who lost her sight completely in a shotgun accident in 2009. She lives with her two children and mother (our participant, A2-P2, a Caucasian office worker), who she frequently asks for help completing various tasks such as grocery shopping, using email and managing her bank account. P2 referred to her loss of vision as "*being constantly in a big black box.*" She also self-identifies

[1] We understood this description could reveal P1's identity. We checked with P1 and he actually preferred this description than a less specific one.

Table 1: The upper part of the table shows the participants with visual impairments (P1-P5), their gender identity, age, marital status, self-described disability or health status, and use of the accessible technologies. The lower part of the table shows the ally participants (A1-P1, A2-P2, A3-P5 are allies of P1, P2, and P5, respectively). For allies, the last two columns represent the social relationship they have with their partner, and the kinds of help they provide (other activities include shopping, navigation, online ordering, household chores, etc), respectively.

| ID | Gender Identity | Age | Marital Status | Self-Described Status | Accessible Tech Use |
|---|---|---|---|---|---|
| P1 | Cisgender Male | 30-40 | Single | Blind | JAWS, NVDA |
| P2 | Cisgender Female | 30-40 | Relationship | Blind, bipolar disorder, learning disability | VoiceOver |
| P3 | Cisgender Male | 80+ | Married | Low vision, physical health condition | Dragon NaturallySpeaking |
| P4 | Cisgender Female | 40-50 | Divorced | Low vision | ZoomText, VoiceOver |
| P5 | Cisgender Male | 60-70 | Married | Blind, hard of hearing | JAWS |
| | | | | **Relationship** | **Assistance Provided** |
| A1-P1 | Cisgender Male | 40-50 | Single | Friend of P1 | Shopping, Navigation |
| A2-P2 | Cisgender Female | 60-70 | Married | Parent of P2 | Banking, Other activities |
| A3-P5 | Cisgender Female | 50-60 | Married | Spouse of P5 | Banking, Other activities |

Table 2: Details of the study sessions for each participant.

| ID | Session 1 (S1 time) | Session 1 (S1 location) | Session 2 (S2 time) | Session 2 (S2 location) | Ally present |
|---|---|---|---|---|---|
| P1 | 8/22/17 10am-4pm | P1 office, apartment | 9/9/17 12-4pm | P1 apartment | Mobile trainer (S1) |
| P2 | 9/28/17 2-7pm | P2 home, friend house, park | 12/3/17 2-6pm | P2 home | A2-P2 (S1, S2) |
| P3 | 12/15/17 8-11am | P3 home, medical office | 4/14/18 10am-12pm | P3 home | None |
| P4 | 2/22/18 11am-3pm | P4 home | 3/13/18 3-5pm | P4 workplace | None |
| P5 | 2/28/18 12-4pm | Campus library, parking lot | 3/17/18 11am-5pm | P5 home, hospital, shop | A3-P5 (S1, S2) |
| A1-P1 | 9/14/17 3-5pm | Workplace of P1 and A1-P1 | 10/1/17 3-5pm | Mall | with P1 (S2) |
| A2-P2 | 9/28/17 6-8pm | P2 home, friend house, park | 12/3/17 2-6pm | P2 home | with P2 (S1, S2) |
| A3-P5 | 2/28/18 12-4pm | Campus library, parking lot | 3/17/18 11am-5pm | P5 home, hospital, shop | with P5 (S1, S2) |

with bipolar disorder and a learning disability. She struggled with using accessible technologies. For instance, she said, "*I can't use JAWS. I just hate that voice, it's so annoying.*"

While prior literature has studied online self-disclosures about stigmatized experiences such as pregnancy loss [5], our study observed whether and why our participants with visual impairments may choose not to disclose their disabilities. For instance, P3 is a retired Caucasian educator with low vision who lives with his wife and asks her for help completing daily tasks. P3 also has a serious health condition requiring him to receive medical treatments several times a week at a local hospital. P3 said he has: *"macular degeneration, continuous loss of the 'center of things;' vision loss starts at the center and then generally progresses through time."* He also said that he struggled with answering a large number of emails due to his low vision, but did not want to broadcast his visual impairment. He explained: "*They are not astute enough to know that I can't read it. And what I am gonna have to do is to contact the people I really want to hear from and tell them that they better call me.*" P3's decision not to broadcast his visual impairment has a practical utility of helping him manage his communications. He only wanted his important contacts to know about his difficulty of viewing emails. As such, he was selectively disclosing his visual impairments.

Prior literature has suggested that using accessible technologies might trigger questions about people's disabilities (e.g., [44]) and make them more vulnerable in public places (e.g., [4]). We found that hiding one's visual impairment is also a way to protect themselves in a public environment. For instance, P4 is a Caucasian disability coordinator with low vision who lives with a housemate and one daughter. She does not drive on her own, and therefore relies on others such as her housemate and co-workers to provide transportation. P4 mentioned that if she needed help to read something in public, she would often hesitate to disclose her visual impairment. Instead, she would say: "*I forgot my glasses at home, can you help read what this is?*" These behaviors may reflect their personal insecurities that can stem from a generally accepted notion of society that they are not part of the mainstream [24].

Similar to prior literature (e.g., [4]), we found such fear of insecurity can also motivate a person with visual impairments to seek help from trusted allies. For instance, P5, a Caucasian retired doctor and a veteran, became blind due to a motorcycle accident. He is also hard of hearing and uses a hearing aid device. He lives with his wife (A3-P5, a Caucasian home-maker) and was heavily dependent on her for grocery shopping or visiting hospitals. He often avoided accepting help from strangers or even friends because of a lack of trust.

## 4.2 Self-Definitions of Privacy

There are many but no agreed-upon definitions of privacy [47]. Instead of defining privacy for our participants, we asked in the exit interview how they would define privacy in their own words. While our participants' privacy conceptualizations were not necessarily new, we note that prior literature rarely covers this aspect for people with visual impairments. Knowing their definitions of privacy can help us unpack their privacy concerns and practices. Three participants (P3, P4, and A1-P1) defined privacy in terms of ownership and control over their personal information. For example, according to P4, privacy meant "*keeping your personal information to yourself; keeping things to yourself that you don't want other people to know.*" Such control over their personal information can also provide them a sense of security. For example, P3 referred to privacy as "*the ability to conduct life confidently and securely, knowing that I will not be surprised by someone telling me things about myself that I have never shared.*"

A1-P1 attempted to define privacy in the context of his "allyship" with P1. He viewed personal information as one's own property and privacy as a right to such property. He explained, "*Things that are private to an individual should remain private unless otherwise stated by the person who owns the rights or the property.*" A1-P1 then gave an example: if P1 asks for help, then he helps but otherwise laundry is his private property and P1 knows how to handle that himself.

Other participants' conceptualizations of privacy focused on "the right to be let alone," a classic definition of privacy proposed by Warren and Brandeis [47]. For instance, P2 felt that privacy could also mean "*alone time privacy.*" Lastly, P5 touched upon the sensitive yet crucial relationship between individuals' privacy and society. According to him, privacy also included, "*acceptance by others of me saying I don't wish to share that info and then in turn, respecting my privacy.*" Overall, these definitions shared a desire of agency and control over one's information and ways of living. With this understanding of our participants' privacy conceptualizations, we now discuss their applications to daily life.

## 4.3 Privacy/Security Concerns and Practices

Similar to the findings of prior work (e.g., [3, 4, 7]), our participants with visual impairments expressed many concerns about privacy and security while using technology on a regular basis, in home, work and public settings.

**Home settings.** We observed some deceptive practices (e.g., scams and malicious software) that plagued our participants with visual impairments. For instance, while conducting the pre-defined tasks portion of our first study session, we asked P1 to demonstrate logging into his email client. During this task, we observed a fake virus warning pop up in his browser. He was unaware of what had happened but just asked, *"what's wrong?"* While any Internet user might click and fall prey to these types of fake warnings, people with visual impairments might be even more vulnerable because they might accidentally click the warning especially if the screen reader does not recognize its existence.

**Work settings.** Our participants were concerned about their private information being inadvertently stolen or compromised on the job, mainly due to enlarged screens or accessibility features leaking private information. For example, during the scenarios portion of P4's study session, she provided us with an instance where she was concerned about shoulder surfing. P4 stated that her work iPad screen is a lot bigger than her phone, thus has the potential for people to easily see her private information. She explained: *"When you have such a big screen, you can't sit there. Most people can do personal stuff, you know, you could see what I'm doing."* She would only check private information in her office.

**Public settings.** Our participants were also concerned about leaking their information in public and adopted various protection strategies, e.g., using earphones while using the screen reader on their phones to check emails, or during ATM withdrawals. However, they also felt it is more challenging for them to hide their information than people without visual impairments. For instance, P4 elaborated during the initial interview that many adept smart phone users without visual impairments can easily check or send a text under the table or otherwise out of sight, but that is not something that she can do. While smartphones have built-in accessibility features (e.g., Android screen reader, iPhone screen curtain to black out the phone screen) for users with visual impairments, they still lack accessible features supporting information hiding.

**Insecurity of information.** Three participants with visual impairments were concerned about insecurity of their information, which is under-reported in prior literature. This is a security concern regarding a potential breach of confidential information [40]. While prior work has shown that people with visual impairments have privacy/security concerns about online transactions (e.g., [3]), P4 was concerned that her data might be transferred from her phone to the cloud, which could be breached. While attempting to complete one of the pre-defined tasks, P4 described a habit of taking photos of her credit cards so that she can enlarge the numbers to see. But she immediately deleted these photos so that they cannot be seen by someone else and would not be in her phone if the device gets lost or stolen. However, she was not sure where the pictures went. She explained: *"..like if it's still in the cloud or somewhere; it worries me."* P4 did not consider this practice as a safe thing to do, because her sensitive financial data might be transferred to and stored in the cloud, where others might gain access to and/or misuse her data. However, she did so to make herself more independent in purchasing. This highlights the trade-offs between independence and privacy/security that people with visual impairments often had to make.

## 5 Social Relationships and Interactions

The everyday practices of people with visual impairments often involve interactions with other people such as allies (e.g., family and friends) and even strangers. This section focuses on these interactions with an attempt to highlight the underlying nature of these relationships and interactions reflecting elements of agency, interdependence, and trust that shape their everyday experiences.

### 5.1 Family Relationships

Family members such as spouses, domestic partners, parents, children or siblings often serve as allies for people with visual impairments because their actions may reflect understanding, commitment, mutual trust, and shared decision making.

**Spouses.** We observed that marriage or domestic partnerships often involve more reliance than friendships and other familial support. Two of our participants with visual impairments (P3 and P5) are both married. For instance, P5 stated during the initial interview that he completely trusts his wife (A3-P5) and is dependent on her to assist him in a wide variety of tasks. P5 recalled an essentiality to share passwords for emails and other critical information about banking with A3-P5. While he was unconscious (resulting from an accident 4.5 years ago when he lost his vision), she needed to access his emails and bank accounts to pay bills and respond to important emails on his behalf, therefore A3-P5 also manages the information being shared between her husband and certain organizations. However, sometimes, organizational policies make it difficult for A3-P5 to help him. She mentioned during the exit interview that she is his power of attorney, which gives her the legal authority to make decisions on his behalf in all financial and legal matters. A3-P5 shared an example: *"There was one situation where somebody said they could not speak to me because their rules were they had to give the information to my husband and literally, he said 'give it to my wife and she will write it down' and he said he could not do it, and he said 'we're basically at an impasse here because basically I can't write it down, so you are going to have to give it to my wife."'* Eventually, they gave the information to her.

The organization in this example practically created inaccessible conditions. P5's interdependent relationship and practices with A3-P5 were in conflict with the organization's policies. While these organizational policies may have been designed to safeguard the security of the users' information, they failed to take an inclusive approach by making the process more accessible for people like P5, who is blind and hard of hearing. These overlapping dimensions of P5's identity might explain why it is difficult for him to write down something said over the phone. While prior work shows that people with visual impairments struggle with password management (e.g., correctly typing passwords [3]), our study presents this novel finding that organizational policies designed to protect users' privacy and security can also be challenging for people with visual impairments such as P5.

**Parents and children.** Most participants with visual impairments often asked their parents and/or children to assist them in different activities such as banking and transportation. For instance, P2 is blind and has bipolar disorder and a learning disability. She often relies on her mother (A2-P2) for help. During the exit interview, P2 explained an incident about finding a person through an online dating site. When she decided to meet this person, her mother offered to drive her to the meeting place. However, her mother did not let P2 get out of the car because she felt the person looked suspicious. In this case, her mother attempted to safeguard her from potential risks but P2 did not really have much control over the situation, because her mother made the judgment for her. This could raise the question of P2's abilities to engage in a relationship independently. P2's multiple disabilities and gender identity could make her particularly vulnerable in these circumstances, which might explain the trade-off her mother made in this case between P2's safety and her social life.

Prior literature suggests that people with visual impairments are concerned about their privacy when asking strangers to help but are comfortable in asking for help from a known person (e.g., [3]). However, a novel finding we observed was that privacy is an important factor when some participants with visual impairments were considering whom to ask for help even if these individuals are their own children. For instance, P4 gave a concrete example in response to a follow-up question we asked her after completion of the entire study of not asking her daughter to fill out forms that require financial information. She explained: *"If I'm filling out a camp scholarship form and my daughter is helping me and it starts asking for salary and blah, blah, blah, I don't. It's mostly because I'm in a divorce situation and I don't want her to accidentally tell her father. Yeah, my kids are pretty trustworthy but I don't trust my ex-husband."* This example highlights how P4 consciously considered potential privacy risks (leaking her salary information to her ex-husband) in mundane activities in her everyday life. While P4 trusted her children to help her, she did not want her ex-husband to know about her financial information, which may put her at risk.

### 5.2 Romantic/Dating Relationships

Romantic/dating relationships can be a source of allyship but might also have privacy risks in this context. For instance, P2 has a boyfriend but avoids to have any other male friends. During the exit interview, she explained: *"It's just if I were to have a guy friend, being blind, you really can't hide that on your phone. Sometimes people can do that. They can hide different things nowadays on phones. I've been told they're able to, but I just find it easier to not even be friends with guys*

*at this point."*

P2 viewed her phone conversations as something private and did not necessarily want others, even those who are close to her, to know about. While visual/aural eavesdropping was reported as a privacy concern in prior work (e.g., [3, 4]), P2's case was different because she was concerned about her boyfriend accessing private/personal data *stored* in her phone. P2 discussed hiding phone conversations from her boyfriend because talking to other male friends might cause misunderstandings between her and her boyfriend. While physical access to a person's phone conversations can happen to anyone, cisgender women with disabilities could be more vulnerable (e.g., in abusive domestic relationships where women are often the victims [22]).

P2 would love to be able to control the visibility of these conversations herself, but she felt the technologies are too complicated for her to learn. As we noted earlier, she also self-identified as having bipolar disorder and a learning disability, struggling with technologies. While technical features such as deleting a phone call record or a text message may allow people to hide certain social interactions on the phone, there are no technical features explicitly marked to "hide conversations on a phone." In order to achieve such goal, one would need to take a socio-technical approach, which may require a good understanding of the phone and its technical features as well as the social implications of using such features. Since P2 generally struggled with technologies, she may find those features overwhelming to learn.

## 5.3 Friends

Friends are also a source of allies for people with visual impairments. For instance, P1 and A1-P1 consider each other to be close friends. During the observation portion of A1-P1's second session accompanying P1, we joined them at the local mall. P1 found something to buy. At the register, there was a line and the cashier was trying to move everyone along quickly. Once P1 reached the cashier, he was prepared to pay with his card. P1 asked A1-P1 to take out his wallet and then A1-P1 identified the card to pay by himself. However, this store only accepted cash. We saw some people in the line looking impatient. Although P1-A1 did not say that he felt pressured in that situation, we observed him taking notice of the people waiting in the line. As a result, A1-P1 voluntarily stepped in, before P1 could ask, and removed cash from P1's wallet to pay the cashier and finish the purchase.

A1-P1 had complete access to his friend's wallet at this point, which in theory could pose financial and privacy risks to P1 (e.g., taking extra cash, knowing how much cash he has, remembering and even misusing his credit card information). In that moment, A1-P1 sensed social pressure (many people waiting in the line), and took control by quickly completing the transaction through cash. By enabling a quicker transaction at the potential cost of P1's financial privacy, A1-P1 prevented P1 from becoming a target of general public frustration by taking too much time to complete tasks that may seem easy for people without disabilities. The inaccessibility of the store was at odds with P1's loss of vision, which influenced his decision to ask his friend to access his wallet and assist with the payment process. However, P1 exercised his agency by consciously making the decision to ask his ally for help. He trusted A1-P1 and was interdependent on him to help perform this financial transaction.

## 5.4 Professional Relationships

Some of our participants with visual impairments also had allies who provided (paid) professional services (e.g., filing taxes). Moreover, they might also ask these allies for help even for tasks falling outside the responsibilities of these allies. For instance, P1 has a mobility trainer to train him with physical navigation. Since P1 is blind, he relies on the assistance of this mobility trainer to develop non-visual cues and landmarks to navigate his physical environment independently. During the observation portion of P1's first session, we observed this mobility trainer assist P1 in finding the route from his apartment to his classroom. As they started from P1's apartment, they stopped by to check his mail on the ground floor of his apartment building. P1 asked his mobility trainer to read his mail. One of the letters seemed to be from a financial institution. P1 asked the mobility trainer to open the letter, who identified it as a check and gave it to P1. We then accompanied them to the bank and a teller recognized P1 and helped him deposit the check. Later we asked P1 for any concerns about other people reading his letters, he explained: *"I am not so much concerned with someone reading my mails, because they are not reading without me telling them. I'm the one asking them, 'can you read this?"'* P1 has developed a trustworthy relationship with the trainer and felt comfortable asking him to read his mail.

There were power dynamics between P1 and the mobility trainer in the form of a professional (paid) relationship between them. Yet these dynamics were diluted by the interplay of an informal friendly relationship, on the basis of which P1 asked his mobility trainer to read his personal mail. Furthermore, by asking his mobility trainer to check his mails, P1 said he has full control over the situation, suggesting that he was exercising his agency. However, at least theoretically, there is a privacy risk for P1 that the mobility trainer could learn/misuse his sensitive information (e.g., bank account and balance). P1's action of giving consent to his mobility trainer to assist him with this task may suggest that either he is potentially aware of the risk involved or may be choosing to ignore such a risk over convenience. Brady and Bigham have discussed people with visual impairments seeking help in a computer-mediated fashion via crowdsourcing or friendsourcing (e.g., asking crowd workers or friends to fix web accessibility issues or to identify the kinds of objects in a pic-

ture) [14]. In comparison, our study shows this phenomenon in an offline, non-technology-mediated manner that could also pose privacy risks. This kind of offline scenario is a rich space for future privacy research and design.

### 5.5 Strangers

People with visual impairments usually do not consider strangers as allies. However, occasionally, our participants still found themselves in situations where they had to ask for help from strangers or were even approached by strangers for help. In those cases, physical safety or security might be a concern. For instance, during the initial interview, P1 described an incident where he tried to walk to a local restaurant alone. At that moment, he was approached by a group of people who offered to escort him to the restaurant. When they all reached the restaurant, this group of people asked for money in exchange for helping him and P1 paid them $10. Later when P1 told A1-P1 about this incident, he said that P1 was not really supposed to give money to people on the street.

### 5.6 Allies' Perspectives

Each of our ally participants also had interesting and unique perspectives about their relationships with those with visual impairments that they often interact with. A1-P1, for example, disclosed during his exit interview that he respected P1's agency and advocated not to offer help unless P1 asks him for such help. He also perceived P1 to be independent and able to manage most aspects of his everyday life very well. However A1-P1 also jumped in to assist him with tasks such as navigating large spaces. A2-P2 took into account not only her daughter's visual impairment but also her mental status, considering how her bipolar disorder and her loss of vision impact her everyday life. In response to a follow-up question we asked A2-P2 after completion of the entire study, she stated: *"I believe in complete right to privacy in all situations dealing with P2. She is an adult and my help to her is strictly for her benefit and I consider any breach of her privacy to be also a breach of trust."* Taking into account P2's multiple disabilities (including visual impairments) as well as her mental condition and gender identity, A2-P2 took privacy concerns very seriously and felt any help and action she would take would only be of her daughter's best interest. The primary priority of A3-P5 was her concern for her husband's safety in general, particularly navigating outside of the home environment. She identified key risk points and assisted him with various tasks in the home. She also felt that her husband being hard of hearing also plays a role in their daily lives.

In summary, the lessons we learned from the perspectives of allies reflect a sense of accommodation and agency. All three allies understood the limitations of our participants' visual impairments along with their multiple identities and provided assistance for them to complete their daily tasks while valuing their privacy and security. They also provided an environment of agency where allies respected our participants' autonomy instead of our visually impaired participants completely depending on their allies for complete assistance.

## 6 Discussion

In this section, we discuss the implications of our results. Following the recommendation of being self-reflexive as researchers [41], we start with information about our research team to contextualize our discussion that follows.

### 6.1 Researcher Self-Disclosure

Our research team consists of individuals across a variety of academic backgrounds and identities. We find this diversity brings strength to our research. Our team has cisgender identified men and women, an Asian American, a Caucasian American, and people from Asian countries. We understand that there are certain privileges associated with our notions of self and we benefit from the ideas of people of color (e.g., [17]). While some of us wear glasses and/or have "hidden" disabilities, they are not the same as visual impairments. Thus, our understandings of the participants might be limited. Nevertheless, we self-identify as allies in the disability community.

### 6.2 Everyday Privacy/Security Practices

While there is a vast body of literature on people's privacy concerns, preferences, and practices (see [23] for a comprehensive review), most empirical work in this line of research focuses on individuals, i.e., the unit of analysis is individuals making decisions about their own privacy. The few exceptions that examine the privacy management and practices of pairs or groups of people tend to focus on social media, for instance, how one user might intentionally or unintentionally disclose information of another person on social media (e.g., [29]).

A crucial element of our approach to study the everyday privacy and security practices of people with visual impairments was to pay close attention to their relationships and interactions with their allies. Therefore, our analysis focuses on not only individual-based but more importantly *group-based* privacy management (i.e. individuals with visual impairments working with their allies). In addition, we have discovered that our participants' self-notions of their disabilities, their conceptualizations of privacy, and the multiple aspects of their identities (e.g., disabilities, gender identity) influenced their privacy and security practices. We elaborate on this discussion by revisiting our two research questions.

**First research question**: *what are the everyday privacy/security concerns, challenges and practices of people with visual impairments in their daily lives?* Our study corroborates with prior literature that shows people with visual

impairments have various physical/offline and online privacy/security concerns such as shoulder surfing and hacking (e.g., [3,4]). Yet our study also provides novel results in terms of how our participants with visual impairments view their disabilities and how their notions of privacy affected their privacy and security practices, for instance, whether to disclose their disabilities (e.g., P3's example of selectively informing his email contacts about his low vision).

Our study also reveals understudied challenges of people with visual impairments in managing their social relationships (e.g., P2's example of meeting someone from an online dating site and her hypothetical example of hiding conversations on her phone so her boyfriend would not misunderstand). P2's attempt to hide conversations on her phone is a form of user appropriation of technologies (e.g., phones).

While user appropriation of technologies for their own purposes has been extensively discussed in the literature (e.g., [49]), we paid extra attention to the complex identities of our participants with visual impairments in understanding their everyday privacy and security practices including technology appropriations. Three participants with visual impairments also have other disabilities or significant health conditions. Some participants are older adults. One female participant was in a divorced situation and another male participant came from an African country with a very different culture. All these aspects of their identities play a role in shaping their varied experiences and privacy/security practices.

Sociologist Erving Goffman has written about disabilities being considered as a stigma by some people and the techniques that people with stigmatized identities used for information control (e.g., covering their identities) [26]. Our study investigated the social/collaborative aspect of privacy and security practices including information control. This leads to our second research question.

**Second research question**: *how do people with visual impairments interact with their allies? What are the privacy or security implications of such interactions?* While prior literature touched on the general phenomenon that people with visually impairments often seek help from their allies (e.g., [3, 4]), our study dove deeper into the social relationships and interactions between adults with visual impairments and their allies, drawing our attention to issues such as agency, interdependence, and trust. We found that our participants interact with their allies in various social settings and everyday activities such as physical navigation, personal finance, doing laundry, grocery shopping, managing social relationships, and using technologies. Our participants sometimes asked their professional service providers to help with tasks that were outside the scope of the service providers' responsibilities (e.g., P1 asking his mobility trainer to check his mails). While this practice might pose a privacy risk, this is understandable because participants with visual impairments have built a trustworthy relationship with their allies.

From our ally participants' perspectives, they respect the independence and privacy of their partners with visual impairments. While these allies often provided help only upon request of their partners, occasionally they acted without explicit request (e.g., P1's friend/co-worker, A1-P1 removing cash from P1's wallet to pay for P1's purchase without his request; and A2-P2 not letting P2 to leave the car to meet with the individual she met on a dating site). These occasional cases highlight the trade-off between respecting their partners' agency and protecting them from unnecessary embarrassment or privacy/security/safety risks. How to better support social interactions and co-decision-making between people with visual impairments and their allies deserves further research.

Furthermore, we observed that our participants with visual impairments were often thoughtful about when to ask whom for what kind of help. Notably, a novel finding of our study is how privacy plays an important role in the decision making of our participants with visual impairments in asking allies for help (e.g., P4's example of not asking her daughter to fill out scholarship forms that ask her salary information, which her ex-husband might then learn). More broadly, our participants with visual impairments hope to achieve more control over their own lives (a form of agency), being able to choose independence or interdependence as they deem appropriate.

## 6.3 Implications for Research and Design

What do these insights of people with visual impairments and their allies mean for privacy research and design?

**Cooperative privacy and security.**
The first key implication is that *privacy management can have an inherently cooperative dimension*. The basic assumption behind most of the existing end-user privacy tools is that privacy management is a personal/individual behavior. Therefore, existing tools are often framed as helping individuals protect their own privacy. However, as our study highlights, people with visual impairments often work closely with their allies to protect their privacy and security. By "cooperative," we intend to call attention to the aspect of mutual assistance in working together towards a common goal in protecting privacy and security. Although prior work [29, 31, 36] discusses the concept of cooperative privacy, our study offers an understudied and nuanced understanding of cooperative privacy practices, which dovetails with the idea of interdependence in the context of people with disabilities and their allies.

In recent scholarship in accessible computing, researchers have highlighted the importance of interdependence. For instance, Bennett et al. advocate interdependence as a frame for research and design of accessible technologies [8]. Traditionally, the main goal of accessible computing has been to support independence of people with disabilities (e.g., independent living). However, Bennett et al. argue that interdependence is also vital, drawing from the literature of disability studies, disability activism, and the social aspects of accessible computing [8]. They use the term interdependence to describe

mutual relations and interactions between people and their environments. Importantly, they highlight that people with disabilities and their allies help each other instead of people with disabilities being "passive recipients of assistance."

Our study results and the concept of cooperative privacy and security align well with Bennett et al.'s framework [2] of interdependence. Our participants with visual impairments and their allies value each other's strengths, differences as well as their holistic characteristics and unique needs to collectively manage their everyday privacy and security. For example, A3-P5's management of her husband's passwords and financial information, A2-P2's accounting for her daughter's multi-faceted disability identity when overseeing her credit cards, bank accounts and physical safety and A1-P1's ability to meaningfully respond to his friend's needs in terms of perceiving potential privacy risks allows each party to fully understand the true strengths and vulnerabilities of one another and use them to their own advantage to provide meaningful feedback, establish a system of trust and communicate potential dangers in relation not only to each other but the world in which they live. Furthermore, each of the allies we studied recognizes the contributions of our participants with visual impairments instead of providing care that does not fully comprehend the strengths of their care recipients. In line with Bennett et al.'s work, these simultaneous and visible relations are key for applying interdependence to cooperative privacy and security because people with disabilities and their allies must delve beyond the surface of providing care and utilize their relationships to truly assess their unique privacy and security needs and experiences.

Privacy and security mechanisms are often focused on the individual's perspective, for instance, a privacy or security warning that a user can act on. In contrast, cooperative privacy fosters interdependence, which is especially beneficial for the every-day privacy management of people with visual impairment. What would a "cooperative" warning look like? Perhaps it could have built-in support for people to seek help or get feedback from others (e.g., allies), for instance, an option on the warning to ask for help. One possible cooperative privacy design could take the form of a mobile app or a website where users with visual impairments could choose to share only with specific allies they invite to the system any information about them, such as schedules and common tasks they perform. If users felt their privacy/security is at risk, they can request help from selected allies, requesting a chat session in real-time where allies would be providing assistance as needed. Users would have full control over the disclosure of any private information that they share via the system. This is just one example of a rich yet largely untapped design space for cooperative privacy and security mechanisms. These types of designs will not only be helpful for people with visual impairments and their allies, but also computer users more generally (e.g., technically savvy users and novice users).

Prior research such as [46] has suggested that group-level analyses of privacy behavior are rare but needed, and our research provides empirical data to support such a claim and a concrete context for exploring group-level analyses. Xu conceptualizes collaborative privacy management as the collaborative strategies and practices that individuals use to protect their privacy as group members [52]. A number of collaborative privacy management tools have been proposed mostly in content sharing scenarios among users of social networking sites [1, 9, 18, 33, 50]. These solutions were usually in a scenario where a person's privacy is violated by another person's behavior, e.g., one person tagging another person in photo posted on social media. None of these mechanisms were designed to cater to people with disabilities and their allies where often people with disabilities face privacy risks and then collaborate with their allies to address those risks.

**Cooperation might introduce risks.** The second key implication has a critical nature: *cooperative privacy management can also introduce new privacy risks in the context of people with visual impairments.* This is especially true when sensitive information related to a person with visual impairments may be exposed to or shared with allies. For instance, when P1 asked his mobility trainer to check his mails, the mobility trainer had access to P1's mails, including financial documents such as a check from the bank. This can be a risk for cooperative privacy in alternative situations because the ally (in this case the mobility trainer) might access personal/sensitive info about people with visual impairments even though this information may be willingly shared by people with visual impairments based on their mutual trust. In theory, the mobility trainer has access to P1's sensitive information, which could be misused. Our study points to the need of designing mechanisms that facilitate these types of cooperative privacy practices while mitigating the potential privacy risks this practice might introduce. This design dimension is crucial for people with visual impairments and can be relevant for other marginalized groups such as children and older adults.

In the accessible computing field, there are a number of proposed tools to support collaboration between people with and without disabilities (e.g., [6, 12, 38, 42, 43, 51]). However, they rarely consider the privacy implications. There are also collaborative systems for people seeking and providing help. For instance, Ahmed et al. designed Suhrid, a collaborative interface for people with low literacy to seek help from helpers [2]. The system only shows the last two digits of a contact person's phone number to mitigate help seekers' concerns about their contacts' privacy [2]. Yet, we are unaware of any collaborative privacy tools that specifically support people with disabilities and their allies.

**Multi-faceted and intersectional identity.** Another key implication is that *when designing privacy/security mech-*

[2] Bennett et al.'s paper is relevant but not a theoretical foundation of our study because it was published after we conducted our study. The findings emerged from our study suggest the importance of interdependence.

*anisms for people from marginalized groups, one needs to pay attention to the multiple and intersecting marginalized identities that these individuals might have*. As seen in our study, many of our participants with visual impairments have multiple marginalized identities (e.g., having multiple disabilities), which are intersecting and thus it can be hard to pinpoint which marginalized identity led to certain experiences or challenges. For instance, P2's challenges with keeping her conversations private on her phone could be influenced by her visual impairments, bipolar disorder and learning disability. It is unclear exactly which and how many factor(s) led to these challenges. Traditionally, privacy research and designs for people with visual impairments tend to focus on a single marginalized identity (i.e., visual impairments). However, privacy designs need to consider the multi-faceted and intersectional aspects of people's marginalized identities. We believe that one practical way to do this is through focus groups and participatory action/design research. Such studies engage these users throughout the design process to learn about their identities and needs.

In addition, the concerns raised by our participants with visual impairments show limitations in current technology designs. Our finding that users with visual impairments cannot easily delete or hide personal communications or information on their devices (or at least the perceptions of such difficulty) implies the need for more accessible solutions. For instance, a device-level, authentication-required feature that hides personal communication across multiple apps such as calls, text messages, and social media posts could be quite useful. In addition, the fact that organizations, policies or systems do not consider enough the needs of people with visual impairments and their allies for handling personal information means that these limitations make life more difficult for all parties involved. Therefore, future designs can explore how to develop more effective and usable mechanisms of sharing information between users with visual impairments and their allies. Existing solutions such as shared accounts in password managers are helpful, but they would not resolve the issue encountered in P5's example. In that particular case, a technical mechanism that allows the organization to add a password directly into P5's password manager would be helpful.

Design processes are often intrinsically power-laden [28]. Arguably, designers are often in a more powerful position than users, in this case, people with visual impairments. To follow a more equitable and empowering approach, results from this study will be fed into our follow-up research where we will collaborate with people with visual impairments and their allies to develop privacy tools using a co-design process.

## 6.4 Limitations

Our study has two sets of limitations. The first is related to our sample. For instance, we had a small sample size, as it was difficult to recruit participants for long hours of participation. Yet, our participants include adults that vary in the spectrum of visual impairments and other diverse identities. The inclusion of allies in our study also allows for some unique relevance to the everyday experiences of participants with visual impairments. Another limitation is that our participants were only recruited in a particular metropolitan area. Similarly, our participants are middle-aged or older adults, thus we do not know much about other age groups. Therefore, these factors limit the generalizability of our results. We also note participants knew our "shadowing" and may have avoided certain behaviors during the study.

The second set of limitations relate to our analysis/interpretation. We focused on participants' own definitions of privacy and security, which made some issues such as physical navigation less relevant for privacy/security even though it was an issue discussed generally. While we tried to consider our participants' various (marginalized) identity dimensions (e.g., disabilities, gender identity) in interpreting their everyday experiences, we lacked data to do an intersectional analysis, which rooted from the lived (subordinate) experiences of Black women and women of color [17]. An intersectional analysis should examine the *interactions* of people's marginalized identity dimensions and how those interactions explain why these people are more marginalized than considering each identity dimension alone (e.g., women of color vs. women or people of color). Our study was not designed this way and we did not intentionally collect data about our participants' complex identities (e.g., social-economic status) and how they interact. Future research should embrace intersectionality more [41].

# 7 Conclusion

We conducted an observational study with interviews in order to gain a deeper understanding of how adults with visual impairments interact with their allies and enact privacy and security in their everyday lives. We paid special attention to our participants' perceptions of their own disabilities, their notions of privacy as well as their social practices. Our findings highlight the need of privacy tools that support cooperative privacy management practices between marginalized users and their allies while mitigating any privacy risks that such cooperation might introduce.

# 8 Acknowledgments

We thank our participants for sharing their insights. We also thank the annoymous reviewers and shepherd for their feedback. This work was supported in part by the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR Grant 90DP0061-01-00) and the National Science Foundation (NSF Grant CNS-1652497).

## References


[1] Shane Ahern, Dean Eckles, Nathaniel S. Good, Simon King, Mor Naaman, and Rahul Nair. Over-exposed?: Privacy Patterns and Considerations in Online and Mobile Photo Sharing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 357–366, New York, NY, USA, 2007. ACM.

[2] Syed Ishtiaque Ahmed, Maruf Hasan Zaber, Mehrab Bin Morshed, Md Habibullah Bin Ismail, Dan Cosley, and Steven J. Jackson. Suhrid: A Collaborative Mobile Phone Interface for Low Literate People. In *Proceedings of the 2015 Annual Symposium on Computing for Development*, pages 95–103. ACM, 2015.

[3] Tousif Ahmed, Roberto Hoyle, Kay Connelly, David Crandall, and Apu Kapadia. Privacy Concerns and Behaviors of People with Visual Impairments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI2015, pages 3523–3532, New York, NY, USA, 2015.

[4] Tousif Ahmed, Patrick Shaffer, Kay Connelly, David Crandall, and Apu Kapadia. Addressing Physical Safety, Security, and Privacy for People with Visual Impairments. In *SOUPS2016*, pages 341–354, 2016.

[5] Nazanin Andalibi and Andrea Forte. Announcing Pregnancy Loss on Facebook: A Decision-Making Framework for Stigmatized Disclosures on Identified Social Network Sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI2018, pages 158:1–158:14, New York, NY, USA, 2018.

[6] Dominique Archambault, Bernhard Stöger, Mario Batusic, Claudia Fahrengruber, and Klaus Miesenberger. A Software Model to Support Collaborative Mathematical Work Between Braille and Sighted Users. In *Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility*, Assets '07, pages 115–122, New York, NY, USA, 2007. ACM.

[7] Shiri Azenkot, Kyle Rector, Richard E. Ladner, and Jacob O. Wobrock. PassChords: Secure Multi-Touch Authentication for Blind People. In *Assets '12 Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility*, pages 159–166, 2012.

[8] Cynthia L. Bennett, Erin Brady, and Stacy M. Branham. Interdependence As a Frame for Assistive Technology Research and Design. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '18, pages 161–173, New York, NY, USA, 2018. ACM.

[9] Andrew Besmer and Heather Richter Lipford. Moving Beyond Untagging: Photo Privacy in a Tagged World. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 1563–1572, New York, NY, USA, 2010. ACM.

[10] Patrick Biernacki and Dan Waldorf. Snowball Sampling: Problems and Techniques of Chain Referral Sampling. *Sociological Methods & Research*, 10(2):141–163, November 1981.

[11] Jeremy Birnholtz and Mckenzie Jones-Rounds. Independence and Interaction: Understanding Seniors' Privacy and Awareness Needs for Aging in Place. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2010*, pages 143–152, 2010.

[12] Jens Bornschein, Denise Prescher, and Gerhard Weber. Collaborative Creation of Digital Tactile Graphics. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*, ASSETS '15, pages 117–126, New York, NY, USA, 2015. ACM.

[13] Richard E. Boyatzis. *Transforming Qualitative Information: Thematic Analysis and Code Development*. SAGE, April 1998.

[14] Erin Brady and Jeffrey P. Bigham. Crowdsourcing Accessibility: Human-Powered Access Technologies. *Foundations and Trends in Human-Computer Interaction*, 8(4):273–372, November 2015.

[15] Erin Brady, Meredith Ringel Morris, and Jeffrey P. Bigham. Gauging Receptiveness to Social Microvolunteering. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 1055–1064, New York, NY, USA, 2015. ACM.

[16] Catherine Courage and Kathy Baxter. *Understanding your users: A practical guide to user requirements methods, tools, and techniques*. Gulf Professional Publishing, 2005.

[17] Kimberle Crenshaw. Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color. *Stanford Law Review*, 43(6):1241–1299, 1991.

[18] Ralf De Wolf, Koen Willaert, and Jo Pierson. Managing privacy boundaries together: Exploring individual and group privacy management strategies in Facebook. *Computers in Human Behavior*, 35:444–454, June 2014.

[19] Bryan Dosono, Jordan Hayes, and Yang Wang. " I'm Stuck !": A Contextual Inquiry of People with Visual Impairments in Authentication. In *Proceedings of the 11th Symposium On Usable Privacy and Security (SOUPS)*, pages 151–168, 2015.

[20] Eilers Denise. Nix the word caregiver. *Dialysis & Transplantation*, 39(5):218–218, May 2010.

[21] Heather A Faucett, Kate E Ringland, Amanda LL Cullen, and Gillian R Hayes. (In)Visibility in disability and assistive technology. *ACM Transactions on Accessible Computing (TACCESS)*, 10(4):14:1–14:17, 2017.

[22] Diana Freed, Jackeline Palmer, Diana Minchala, Karen Levy, Thomas Ristenpart, and Nicola Dell. "A Stalker's Paradise": How Intimate Partner Abusers Exploit Technology. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 667:1–667:13, New York, NY, USA, 2018. ACM.

[23] Simson Garfinkel and Heather Richter Lipford. *Usable Security: History, Themes, and Challenges*. Morgan & Claypool Publishers, 2014.

[24] Rosemarie Garland-Thomson. Integrating Disability, Transforming Feminist Theory. *NWSA Journal*, 14(3):1–32, 2002.

[25] Barney Glaser. *Theoretical sensitivity*. The Sociology Press, 1978.

[26] Erving Goffman. *Stigma: Notes on the Management of Spoiled Identity*. Touchstone, New York, reissue edition, June 1986.

[27] Simon Hayhoe and Azizah Rajab. Ethical considerations of conducting ethnographic research in visually impaired communities. In *The European Conference on Educational Research*, Oxford, September 2000.

[28] Lilly Irani, Janet Vertesi, Paul Dourish, Kavita Philip, and Rebecca E. Grinter. Postcolonial Computing: A Lens on Design and Development. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 1311–1320, New York, NY, USA, 2010. ACM.

[29] Haiyan Jia and Heng Xu. Autonomous and interdependent: Collaborative privacy management on social networking sites. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 4286–4297. ACM, 2016.

[30] Shaun K. Kane, Chandrika Jayant, Jacob O. Wobbrock, and Richard E. Ladner. Freedom to Roam: A Study of Mobile Device Adoption and Accessibility for People with Visual and Motor Disabilities. In *ASSETS'09, October 25–28, 2009, Pittsburgh, Pennsylvania, USA*, pages 115–122, 2009.

[31] Jan Kolter, Thomas Kernchen, and Günther Pernul. Collaborative privacy–a community-based privacy infrastructure. In *IFIP International Information Security Conference*, pages 226–236. Springer, 2009.

[32] Teppo Kroger. Care research and disability studies: Nothing in common? *Critical Social Policy*, 29(3):398–420, August 2009.

[33] Airi Lampinen, Vilma Lehtinen, Asko Lehmuskallio, and Sakari Tamminen. We're in It Together: Interpersonal Management of Disclosure in Social Network Services. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 3217–3226, New York, NY, USA, 2011. ACM.

[34] Gloria Mark, Victor M. Gonzalez, and Justin Harris. No Task Left Behind?: Examining the Nature of Fragmented Work. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, pages 321–330, New York, NY, USA, 2005. ACM.

[35] Mari Mikkola. Feminist Perspectives on Sex and Gender. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2017 edition, 2017.

[36] Darakhshan J Mir, Yan Shvartzshnaider, and Mark Latonero. It takes a village: A community based participatory framework for privacy design. In *2018 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 112–115. IEEE, 2018.

[37] M Naftali and L Findlater. Accessibility in context: understanding the truly mobile experience of smartphone users with motor impairments. In *Proceedings of the 16th international ACM SIGACCESS conference on Computers and accessibility - ASSETS 2014*, pages 209–216, 2014.

[38] Shotaro Omori and Ikuko Eguchi Yairi. Collaborative Music Application for Visually Impaired People with Tangible Objects on Table. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '13, pages 42:1–42:2, New York, NY, USA, 2013. ACM.

[39] Nektarios Paisios, Alex Rubinsteyn, and Lakshminarayanan Subramanian. Exchanging cash with no fear: A fast mobile money reader for the blind. *Frontiers of Accessibility for Pervasive Computing*, June 2012.

[40] Chad Perrin. The CIA Triad. *TechRepublic*, 2008.

[41] Ari Schlesinger, W. Keith Edwards, and Rebecca E. Grinter. Intersectional HCI: Engaging Identity Through Gender, Race, and Class. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 5412–5427, New York, NY, USA, 2017. ACM.

[42] John G. Schoeberlein and Yuanqiong Wang. Accessible Collaborative Writing for Persons Who Are Blind: A Usability Study. In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '12, pages 267–268, New York, NY, USA, 2012. ACM.

[43] Jonathan Schull. An Extensible, Scalable Browser-based Architecture for Synchronous and Asynchronous Communication and Collaboration Systems for Deaf and Hearing Individuals. In *Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility*, Assets '06, pages 285–286, New York, NY, USA, 2006. ACM.

[44] Kristen Shinohara and Jacob O. Wobbrock. In The Shadow of Misperception: In the Shadow of Misperception: Assistive Technology Use in Social Interactions. In *Proceedings of the 29th Annual Conference on Human Factors in computing systems - CHI 2011*, pages 705–714, May 2011.

[45] Tobin Siebers. *Disability Aesthetics*. University of Michigan Press, 2010.

[46] H. Jeff Smith, Tamara Dinev, and Heng Xu. Information Privacy Research: An Interdisciplinary Review. *MIS Quarterly*, 35(4):989–1016, December 2011.

[47] Daniel Solove. A Taxonomy of Privacy. *University of Pennsylvania Law Review*, 154(3):477–564, 2006.

[48] Norman Makoto Su and Gloria Mark. Communication Chains and Multitasking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 83–92, New York, NY, USA, 2008. ACM.

[49] Pierre Tchounikine. Designing for Appropriation: A Theoretical Account. *Human–Computer Interaction*, 32(4):155–195, July 2017.

[50] Na Wang, Jens Grossklags, and Heng Xu. An Online Experiment of Privacy Authorization Dialogues for Social Applications. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, CSCW '13, pages 261–272, New York, NY, USA, 2013. ACM.

[51] Mike Wu, Ronald M. Baecker, and Brian Richards. Field Evaluation of a Collaborative Memory Aid for Persons with Amnesia and Their Family Members. In *Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '10, pages 51–58, New York, NY, USA, 2010. ACM.

[52] Heng Xu. Reframing Privacy 2.0 in Online Social Networks. *University of Pennsylvania Journal of Constitutional Law*, 14(4):1077–1102, March 2012.

[53] Hanlu Ye, Meethu Malu, Uran Oh, and Leah Findlater. Current and Future Mobile and Wearable Device Use by People with Visual Impairments. In *Proceedings of the 32nd Annual ACM conference on Human factors in computing systems - CHI 2014*, pages 3123–3132, April 2014.

## Study Script

**Initial interview**

1. What is your age?
2. What is your self-identified gender?
3. What is your occupation?
4. Who do you live with?
5. Walk me through what an average day is like for you. (Both weekday and weekend)
6. How would you self-describe your visual abilities? How would you self-describe your disability status more broadly? (encourage them to be as specific as they feel comfortable)
    a. (If participant mentions having visual impairment or need for visual aids) How long have you had this impairment/used visual aids?

(*For ally participants*)

7. How do you know [insert name of care recipient]?
8. How long have you known [insert name of care recipient]?
9. How often do you assist [insert name of care recipient] during an average day?
10. With which tasks do you typically help [insert name of care recipient]?
11. What information about [the care recipient] do you need to provide the help?
12. What do you typically use the Internet for?
13. How often do you browse the Internet for news or general information? How often do you check your email? How often do you shop online?
14. Do you use social media? How often do you check those accounts?
15. Do you use online banking? How often do you check your accounts online?
16. Do you have a personal computer? What is the model?
    a. Take note of operating system.
    b. Is there any reason you chose that device over other options?
    c. Which browser do you use most on your personal computer?
17. Do you have a mobile phone? What is the model?
    a. Take note of operating system.
    b. Is there any reason you chose that device over other options?
    c. Which browser do you use most on your mobile phone?
18. Do you use any other Internet-connected technologies?
19. Have you ever encountered any difficulties or challenges using technologies? Could you give me a concrete example?
    a. (If prompting is needed) When was the last time you had such an experience? Please describe it.
21. Do you have any concerns using these technologies? Could you give me a concrete example?
    a. (If prompting is needed) When was the last time you had such an experience? Please describe it.
22. Do you use any accessibility features or devices on your phone and computer?
23. Do you share your devices with other people? Who?
24. When you use your mobile phone, laptop, or accessibility devices, do you use them by yourself or in the presence of others?
25. Approximately how often do you ask for help with technology from another person?

**Observation**

Activity Entries
Take note of the interviewee 's (work / home) environment early in the shadowing process.
**For each activity, take note of the following:**

1. Start/end times.
2. Where does the activity take place (which specific room)?
3. Who is in the vicinity for this activity?
4. What personal or sensitive information is relevant to completion of this activity?
5. Is a commodity device (such as a smartphone or laptop) involved? If so:
    i. What type of device?
    ii. What operating system?
6. Is an assistive technology (such as a screen reader or magnifier) involved? If so:
    i. What type of device?
    ii. Does the assistive technology fully solve the accessibility problem for this activity? If not, record the shortcomings.
7. Are other people involved? If so:
    i. What is their relationship to the participant?
    ii. Are they an ally and/or the primary point of contact for help?
    iii. Do they have a disability?
    iv. What information does the participant provide to the other person?
    v. Does the activity require other people to be involved?
    vi. Take note of any hesitation the participant expresses in seeking help or divulging information.
8. (*For ally participants*) Are they trying to help the person for which they serve as an ally?
    i. What is the person trying to do?
    ii. Did that person ask for help? If so, what sort of help are they asking for?
    iii. What help does the participant provide?
    iv. What information did the person they're helping relay to the ally?
    v. What information does the ally access in the course of helping this person (beyond what they learned from the person they're helping)?
    vi. Does the ally do anything else with the information once they have finished with helping?
9. Does the participant use an offline method for this activity when an online method is available?
10. Do any privacy challenges arise?
    i. Does the participant try to address this challenge? If so, how?
11. Do any usability challenges arise?
    i. Does the participant try to address this challenge? If so, how?
12. Do any accessibility challenges arise?
    i. Does the participant try to address this challenge? If so, how?

Additional Tasks and Scenarios
**Tasks:** Now I'd like to observe how you perform specific activities with your personal devices. As I mentioned before, please feel free to say no if you feel uncomfortable with any task.

1. (*If participant uses assistive technology*) Can you walk me through the features of [insert assistive technology] that are most relevant to your daily tasks?
2. (*Ask for permission*) Can you show me how you check your email on…
    i. Your desktop (or laptop)?
    ii. Do you check your email on your phone? If so, can you show me how you do it?
        a. Which device do you use more frequently for this activity?
    iii. Do you ever experience any difficulties or challenges with checking your email?
    iv. What's your strategy for remembering your email account password?
3. (*Ask for permission*) Can you show me how you check your social media accounts on…
    i. Your desktop (or laptop)?
    ii. Do you also check your social media accounts on your phone? If so, can you show me how you do that?
        a. Which device do you use more frequently for this activity?
    iii. What do you like to check on your social media accounts?
    iv. Do you ever experience any difficulties or challenges with checking these accounts?
    v. Does your strategy for remembering social media passwords differ from the way you remember your email password? If so, how?
4. (*Ask for permission*) Can you show me how you check your financial/bank accounts on…
    i. Your desktop (or laptop)?
    ii. Do you also check your financial/bank accounts on your phone? If so, can you show me how you do that?
        a. Which device do you use more frequently for this activity?
    iii. Do you ever experience any difficulties or challenges with checking these accounts?
    iv. Does your strategy for remembering financial/bank passwords differ from the way you remember your other passwords? If so, how?

**Scenarios:** Next, I want you to imagine yourself in each of the following scenarios and tell me what challenges or concerns may arise for you. If you've had such an experience before, describe to me the most recent time it happened.

1. You need to run some errands around town, such as buying groceries and going to the post office. (*If needed, prompt the participant to talk about how they get around town*.)
2. (*If applicable*) You schedule a pickup with Call-A-Bus.
3. You share your medical history with an assistant in the waiting room at the doctor's office.
4. You read your email at a bus stop, and several other people are waiting or passing by.
5. You type your email password into your phone in the breakroom at work.
6. You withdraw cash from an ATM.

(Only for ally participants)
**Tasks:** Earlier today, you mentioned that you typically help [insert name of care recipient] with [*list tasks from initial interview for which the ally provides help*]. If you don't mind, I'd like to walk through how you provide that help. As I mentioned before, please feel free to say no if you feel uncomfortable with demonstrating any of these things.

*For each task in the list from the initial interview, ask the ally to demonstrate how they provide help.*
*Ask the following questions during each walkthrough*:

1. Do you typically ask [insert name of care recipient] whether they need help in this task, or do you provide help unprompted?
2. Do you always feel prepared to help in this task? If not, describe the last time you felt unprepared to help with the task.
3. Do you have to ask [insert name of care recipient] for specific information when you complete this task? What information?
4. Is there any personal information relevant to this task that you already know? What information is that?
5. Do you ever feel uncomfortable providing help with this task? Which components of the task make you uncomfortable? Why?
6. Do you ever have trouble helping [insert name of care recipient] with this task? Please describe the specific challenges for me.
7. How do you overcome these challenges?
8. Have you ever made mistakes with this task in the past? If so, what have been the consequences?
9. Have you ever regretted providing certain help? Could you give me a concrete example?
10. Do you ever require the help of others beyond [insert name of care recipient] to complete this task?

*Additionally, take note of the following:*

1. Does the ally interact with the care recipient when providing help, or do they complete the entire task independent of the care recipient?
2. Do they access more information than they need for the task? Was this a mistake?

**Exit Interview**
(*For participants with visual impairments*)

1. Does your use of [insert assistive technology] change between the home and work/public environment? If so, why?
2. Do you ever hesitate to ask for help from your ally? If so, under what circumstances?
3. Beyond your ally, who do you trust to help you with tasks that are inconvenient?
4. Do you ever hesitate to ask for help from strangers? If so, under what circumstances?
5. Have you ever regretted asking someone for help? Or provide information to someone so that they can help you? Can you give me a concrete example?

(*For ally participants*)

6. Do you ever provide help to [insert name of care recipient] unprompted? Can you give me concrete examples?
7. Do you ever have trouble helping [insert name of care recipient]? Can you give me concrete examples?
    - How do you overcome those challenges?
8. Do you ever feel uncomfortable with the type of information you handle when providing help? Can you give me concrete examples?
9. Do you ever access or see more information from [insert name of care recipient] than the task requires? Can you give me concrete examples?

10. In what locations do you feel comfortable using your…
    i. Email accounts?
    ii. Social media accounts?
    iii. Financial accounts?
11. What does privacy mean to you?
12. What privacy concerns do you have when browsing on the Internet?
13. What privacy concerns do you have beyond Internet browsing?
14. Do you have specific privacy concerns about your mobile phone, personal computer, or work-related devices?
15. Do you ever check out online advertisements? Have you come across any advertisements that appear to be tailored to you? Can you provide a concrete example?
16. How do you currently cope with the privacy concerns or challenges you experience?
17. Do you have suggestions for solutions to these privacy challenges?

**Wrap up**

We really appreciate all the time you've given us. As we wrap up, let me summarize some of the key points I've learned today.

1. Create a large interpretation of your learning about the user's daily activities. The wrap-up is an opportunity to summarize what you learned about the user's experience. It is a way for you to check your high-level understanding with the user. Specifically mention the following:
    i. Computer usage
    ii. Online activities
    iii. General use of technology
    iv. Major challenges encountered during an average day
    v. Privacy attitudes, concerns, and needs
    vi. Participant suggestions for protecting individual privacy
2. Clear up any thought processes or observations that need further clarification.
3. Ask the participant to reflect on their experience with the observation, and ask whether there is anything else in terms of privacy, usability, or accessibility they would like to add.
4. Ask the participant after both sessions: **"**Was there anything that made you uncomfortable today?"
5. Ask the participant after the first session: "Is there anything I should do differently when observing you?"
6. Can the user suggest another interested person with visual impairments who would like to get involved with the study?
7. Thank the user for his/her time and give the user their compensation. Exchange contact information so that the user/researcher can ask any follow up with any questions.

# Privacy and Security Threat Models and Mitigation Strategies of Older Adults

Alisa Frik,[1,2] Leysan Nurgalieva,[3] Julia Bernd,[1] Joyce S. Lee,[2] Florian Schaub,[4] Serge Egelman[1,2]
[1]*International Computer Science Institute (ICSI)*
[2]*University of California, Berkeley*
[3]*University of Trento*
[4]*University of Michigan*
*afrik@icsi.berkeley.edu, leysan.nurgalieva@unitn.it, jbernd@icsi.berkeley.edu,*
*joyce@ischool.berkeley.edu, fschaub@umich.edu, egelman@cs.berkeley.edu*

## Abstract

Older adults (65+) are becoming primary users of emerging smart systems, especially in health care. However, these technologies are often not designed for older users and can pose serious privacy and security concerns due to their novelty, complexity, and propensity to collect and communicate vast amounts of sensitive information. Efforts to address such concerns must build on an in-depth understanding of older adults' perceptions and preferences about data privacy and security for these technologies, and accounting for variance in physical and cognitive abilities. In semi-structured interviews with 46 older adults, we identified a range of complex privacy and security attitudes and needs specific to this population, along with common threat models, misconceptions, and mitigation strategies. Our work adds depth to current models of how older adults' limited technical knowledge, experience, and age-related declines in ability amplify vulnerability to certain risks; we found that health, living situation, and finances play a notable role as well. We also found that older adults often experience usability issues or technical uncertainties in mitigating those risks—and that managing privacy and security concerns frequently consists of limiting or avoiding technology use. We recommend educational approaches and usable technical protections that build on seniors' preferences.

## 1 Introduction

Due to increasing life expectancy, the number of people in the U.S. over 65 is expected to double by 2060 [79]. The need for professional care is rising accordingly, while the labor market for caregivers is projected to shrink [59]. These factors are stimulating investment in emerging "smart" technologies for older adults—aimed at sustaining independent living, increasing quality of life, and mitigating health issues via early detection [83]. Emerging smart technologies such as wearable medical devices, fall sensors, and therapeutic robots [10] may yield benefits, but due to their novelty, complexity, and propensity to collect vast amounts of information, they also pose security and privacy risks.

Due to limited technological literacy and experience, and because of declining physical and mental abilities [44, 96], older adults are particularly unaware of and susceptible to those privacy and security risks [5, 16]. Specifically, older adults have less knowledge of Internet security hazards [36, 40], use technology less frequently [19, 28, 40, 43, 52, 101], are more vulnerable to security risks [41], and are more often targeted for attacks [48] than younger populations. Lack of security knowledge and experience generally correlates with riskier behaviors [45, 71, 73]. Indeed, older adults seem generally less likely to protect against privacy and security risks [57, 62, 85, 99, 101]—though the subject of older adults' privacy management has not been investigated comprehensively.

While seniors often express privacy and (to a lesser extent) security concerns in relation to technology [35, 64, 87], their views are underrepresented in privacy and security research. At the same time, the limited literature on the topic shows that privacy preferences of older adults are heterogeneous [36] and fine-grained [16, 47], and thus warrant further exploration.

The goal of our research is to inform the design of effective systems that empower older adults to make informed decisions, to have better control over their personal data, and to maintain better security practices. To this end, we conducted semi-structured interviews with 46 older adults (65–95 years old). We identify their common security and privacy concerns and threat models, behaviors and strategies to mitigate perceived risks, usability issues with current protections, learning and troubleshooting approaches, and misconceptions regarding security and privacy.

We add depth to current models of how older adults'


*USENIX Symposium on Usable Privacy and Security (SOUPS)* 2019.
August 11–13, 2019, Santa Clara, CA, USA.

relatively low technical knowledge and experience and age-related declines in abilities amplify their vulnerability to certain risks, and found that health and living situations and financial considerations also play an important role. We also found that older adults often experience usability issues or technical uncertainties in mitigating those risks—and that managing privacy and security concerns frequently entails them limiting or simply avoiding use of new technologies. Based on the identified preferences of older adults, we offer privacy- and security-enhancing recommendations for product developers and for educational efforts.

## 2 Related Work

Technological solutions aiming to meet older adults' needs span different domains (e.g., health, nutrition, safety, or navigation [10]) and forms (e.g., wearable, ambient, or camera-based devices [98]). Both aspects factor into what data is collected: wearable devices, for instance, enable collection of orientation, movement, and vital signs with embedded gyroscopes, accelerometers, and other sensors [56, 92]. Context-aware systems use sensors as well, often with the addition of image capture, computer vision, and artificial intelligence to monitor activities or to detect anomalies [e.g., 20, 31]. Likewise, dynamic care robots [e.g., 3, 51, 74] leverage sensors and sometimes cameras for medication management or companionship. Many emerging technologies are connected via Wi-Fi, Zigbee, or similar protocols [e.g., 77, 86], integrating wearable devices with context-aware sensors into a larger ecosystem.

The effectiveness and quality of assistance in critical situations often rely on collecting extensive data. However, extensive monitoring and surveillance trigger privacy and security concerns among users of such technologies [35, 64, 87].

Older adults' privacy concerns and risk perceptions are often different from the concerns of the better-studied younger population [34, 36]. Trust has been identified as a core factor affecting older adults' adoption of ubiquitous computing technologies [22, 24, 65]. However, Knowles and Hanson [54] found that the language of (dis)trust was more relevant to larger value-related issues around digital technologies than to practical decision-making about adoption.

Knowles and Hanson therefore argue that technology adoption should not be viewed as indicating trust or acceptability [54]. Seniors' concerns about monitoring systems include invisible audiences, and absence of feedback when systems are in use or when data is accessed [92]. Other research suggests that some seniors are concerned about who accesses data, how often, and at what level of detail [4, 47, 49]. Although older adults tend to rely on family members in "dealing with technology" [47, 75], delegation of security choices should not be considered a safe behavioral strategy [32].Additionally, older adults may have misperceptions about security, for example, due to over-reliance on surface cues and affordances[e.g., 47].

On the other hand, misconceptions about data collection may raise false concerns that can be mitigated by appropriate explanations [97]. Older adults are also capable of using data controls and security strategies in certain cases, such as basic password encryption [14, 47]. Furthermore, individual differences are found to heavily affect privacy and security preferences: seniors with severe health conditions are more likely to share their information [11, 97] and generally value independence and safety more than privacy [26, 27, 67].

Seniors also represent a more heterogeneous population than younger people [39, 60], due to differences in their health conditions, education, living conditions, and experience. Physical and cognitive impairments may further complicate usability issues. These findings suggest that older adults' privacy and security attitudes and mental models are context-dependent and heterogeneous in nature.

## 3 Methods

We conducted 1–1.5 hour semi-structured in-person interviews, in which we discussed: (1) privacy- and security-related concerns and threats and (2) risk management strategies.[1]

We reached out to inhabitants of nursing homes and senior residences, members of senior centers, and organizations for retired people in the San Francisco Bay Area. We screened potential participants using surveys in several formats—online, phone, paper, and in person—but excluded individuals with serious cognitive impairments and non-English speakers. With IRB approval, we conducted interviews in May–June 2018 with 46 participants at their residences or at public senior centers (their choice). We paid $20 as compensation. We administered exit surveys about participants' individual characteristics.

The structure of our interviews was inspired by Zeng et al. [100], who interviewed 15 smart home inhabitants about their privacy and security attitudes and behaviors. However, our study discussed healthcare and wearable devices in addition to context-aware smart technologies, and involved both users and *non*-users of such technologies.

We audio recorded the interviews and had them professionally transcribed. Three researchers iteratively developed a codebook by independently coding subsets of transcripts and jointly resolving conflicting codes. To maximize the value of thematic analysis, 4 researchers used a holistic coding approach, in which at least 2 coders coded each entire interview, independently selecting excerpts to annotate. All 4 coders then resolved disagreements at the interview level (so that at least 3 out of 4 agreed).

**Limitations.** We conducted our study in an urban/suburban area with relatively good technology resources, programming,

[1] The interview guide—which also includes questions that will be explored in later papers—can be accessed at https://blues.cs.berkeley.edu/wp-content/uploads/2019/06/Interview-guide.pdf. Entry and exit survey instruments can be accessed at https://blues.cs.berkeley.edu/wp-content/uploads/2019/06/Survey-Instruments.pdf.

and services for older adults, and a relatively high average income due to the high cost of living. Our sample is therefore not fully representative, though it is diverse in terms of level of independence, health, living arrangements, and activity. Because we primarily recruited through senior centers, programs, and living facilities, which often offer computer classes, our participants may be more likely to have attended or at least heard about such classes, and therefore may have more awareness of privacy and security issues. Finally, some participants may have experienced interview fatigue.

## 4 Participants

Our 46 participants are 65–95 years old (mean=76), 65% female, mainly white (76%), with self-reported native or bilingual English proficiency (45%) or advanced non-native proficiency (37%). They are diverse in terms of income, health, and care situations (Table 1). The majority have an advanced (44%) or Bachelor's (33%) degree. The majority live alone (63%).

| Individual characteristics | N | % |
|---|---|---|
| **Income level** | | |
| Less than $35,000 | 16 | 35% |
| $35,001-75,000 | 16 | 35% |
| $75,001-150,000 | 6 | 13% |
| More than $150,000 | 4 | 9% |
| Preferred not to answer | 4 | 9% |
| **Housing** | | |
| Independent/assisted living (w/ health facilities) | 6 | 13% |
| Senior/retirement community | 10 | 22% |
| Mainstream housing (rent or own) | 30 | 65% |
| **Self-reported health conditions** | | |
| Excellent | 8 | 17% |
| Good | 23 | 50% |
| Fair | 11 | 24% |
| Poor | 3 | 7% |
| Very poor | 1 | 2% |
| **Caregivers** | | |
| No one | 37 | 80% |
| Hired caregiver | 4 | 9% |
| Informal caregiver | 3 | 7% |
| Both hired and informal caregivers | 2 | 4% |

Table 1: Participant characteristics based on survey responses.

Table 2 shows usage of common devices (11% use none of these). For comparison, 78% of the US general adult population use computers daily or sometimes [7], and 36% use all three [6].

Figure 1 shows participants' self-reported facility with performing certain tasks.[2] Most found basic tasks very easy or somewhat easy. For more advanced tasks, they were more likely to say they had never tried them than to rate them as difficult.

[2]Percentages are out of 45; one participant skipped this question.

| Device Type | Daily | Sometimes | Never |
|---|---|---|---|
| Mobile phone, smartphone | 52% | 22% | 26% |
| Tablet | 22% | 24% | 54% |
| Computer/laptop | 61% | 22% | 17% |
| All three | 11% | 39% | – |

Table 2: Device use among participants.

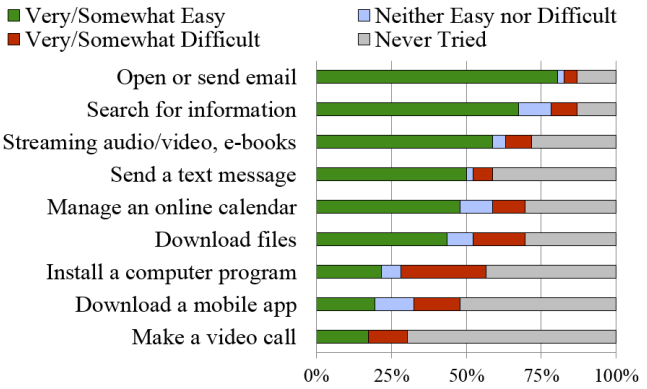



Figure 1: Participants' facility at performing online tasks.

## 5 Findings

In our interviews, we identified privacy and security concerns (§5.1); mitigation and learning strategies to alleviate privacy and security risks, as well as usability issues with those mitigations (§5.2), and misconceptions about data practices (§5.3). In general, the threat models and associated misconceptions that came up in our interviews are also common among the younger population [cf. 40, 62, 95]. However, we found that, due to infrequent use of technology and limited technical knowledge, health and living situations, financial considerations, and age-related ability declines, older adults may be particularly vulnerable to certain risks, and face more issues with mitigating them.

### 5.1 Privacy and Security Threat Models

In this subsection, we describe participants' models of perceived privacy and security threats, and discuss how older adults may be particularly vulnerable to certain risks. We found that participants are concerned about the opaqueness of data flows, especially in emerging and unfamiliar technologies. Even if they do not engage with such technologies directly, some still feel exposed to the privacy and security threats those technologies pose (e.g., passive data collection). Privacy choice is particularly limited for residents of senior care facilities.

#### 5.1.1 Taxonomy of Threat Models

Our participants' privacy and security threat models can be categorized in terms of the *activities* that can lead to security and privacy risks, along with the *consequences* of

privacy and security violations. The discussion below follows Solove's taxonomy [81] in dividing harmful activities into 4 types: information collection, information processing, information dissemination, and privacy invasion. We include a comprehensive breakdown of participants' identified threats and concerns, in terms of both harmful activities and harmful consequences, in Appendix A.

**Information Collection.** One major concern is the lack of transparency about information gathering and people's inability to control it. The issue was raised by 28 out of 46 participants, including 26 who specifically mentioned concerns about collection of data without meaningful notice and consent.

Existing literature has documented the general lack of effective consent mechanisms and transparency regarding data collection practices [78]. This concern is amplified among older adults due to lower technical literacy and experience [85]. For instance, synchronization across devices is a "black box" and source of concern for some participants. Even a participant who volunteers helping others configure devices, considered a computer expert by peers, has trouble tracking it: *"I was concerned that [...] you think you know what shares, but stuff can wind up on another computer so easy with an Apple." "The sharing just surprises me sometimes. You don't know how stuff can go from one to the other, you are surprised it's there."* (P123).

One participant noted that, although data collection by corporations is not new, the Internet and related technologies make collection processes easier, more ubiquitous, and at the same time more opaque: *"The old way, it seemed there was an appearance of consent. [...] Now it's just more seamless,"* (P71).

The inability to control passive audio and video collection by phones and computers—and especially by emerging technologies, like smart TVs, fall detectors, voice assistants, and home-control systems—is of specific concern for 17 participants. Participants believe that information collected by such means may be used for unsolicited marketing, perpetrating physical harm, or violating personal privacy (*"It's scary. Just like, it invades—if the government were to put a microphone in everybody's house and listen to everything you say, people would object. But they are voluntarily putting these devices in their homes and it's doing the same thing,"* P108).

A less common, yet still important, concern participants voiced was about their privacy as bystanders. These concerns were most often related to emerging technologies, such as voice-activated, video-monitoring, and other context-aware systems. Older adults may not be familiar with smart systems, or may even deliberately avoid using them, but nonetheless they are often exposed to data collection by such devices. They may not know how to recognize when smart systems are in use, and may feel uncomfortable about their use by others. This discomfort can contribute to general feelings of helplessness about maintaining control over data collection in the age of ubiquitous computing (*"All my charge cards, all my whatever, everybody knows exactly what I'm doing, even though I never put it on a computer. It's on a computer from someplace else. [...] Every phone call you make is recorded somewhere,"* P43).

Participants often personified data collection processes as though they are conducted by individuals (even when they know it is automated) (*"Whenever you look something up, you get an ad. So a lot of people are reading what you do,"* P5). In some cases, they attributed responsibility for those processes directly to top management: *"The computer [...] probably tracks what you are watching, what you are going to, what you are inquiring about, and keeps a record of it internally. [Interviewer: For what reason?] Because Steve Jobs made it that way. To track data,"* (P69); *"On Facebook, I started—and then they have this Zuckerberg thing about what they were capturing,"* (P104).

***Information collection in senior care.*** Surveillance is another common data collection concern, mentioned by 20 of 46 participants. While a few participants raised broad concerns about government and political surveillance, or referred to personal stalking, the most prominent form of surveillance discussed by older adults was "care surveillance" [30, 68]. Monitoring of older adults by family members, medical staff, or facility management is usually initiated for benign purposes (e.g., to track health status and well-being, or to determine the appropriate level of care). However, such surveillance still induces anxiety, annoyance, and privacy concerns among our elderly participants (*"I know a lot of these devices have cameras in them, and rightly so because they are designed to be helpful, but you know, it's always a concern, I think, when you are using some of the new electronic, is how private are the things that you do,"* P22).

Surveillance is especially common in assisted living facilities or nursing homes. Senior care facilities try to maximize their quality of care and ensure safety while minimizing staff (*"There are sensors so that if you don't go up and go to the bathroom, someone will come down the hall and see if you are okay,"* (P69); *"If [my wife] goes out the front door, it activates a buzzer. There are other residents there who have the same device. [...] The ones that are considered [...] 'exit-seeking,' "* P15). The use of surveillance in care facilities may also be driven by accountability and liability reasons, such as contractual and legal obligations, or to review staff responses to incidents.

Moving to a care facility is often motivated by deteriorating health conditions and the need for a higher level of care. Therefore, older adults living in care facilities are often resigned to giving up privacy in exchange for safety and care (*"You cede a lot of your personal privacy rights when you move into a place like this, in exchange for services being rendered to you. So I think that's a different kind of a setting than somebody that is living in a private setting and would be using devices,"* P71). This finding is consistent with prior literature about tradeoffs between privacy and quality of care [15, 53], and with studies showing positive correlations between the acceptance of privacy risks and deteriorating health conditions [23].

Surveillance is also a concern among seniors who live independently. On the one hand, home care surveillance can

prolong independent living [review in 76]. On the other hand, home care surveillance limits older adults' independence and privacy. Seniors who live independently, and want to protect both health and independence, recognize this tradeoff as a decision they will have to make, if they need more care in the future. As seen in previous research [89], seniors are concerned about how they will balance privacy concerns with the benefits of care surveillance in preserving their autonomy (*"I would probably choose [a wall sensor that detects] presence over having to share a room with somebody being in a nursing home. So if I could stay in my own abode [...] that is a concession that I would make,"* P24).

**Information Processing.** Almost half of our participants (19 of 46) mentioned aggregation of personal information about individuals from multiple sources, such as web browsing records, smart TVs, and wearable fitness trackers. While some participants find customized recommendations beneficial, most find individual profiling concerning (in some cases both).

However, a few participants showed limited understanding of how inferences can be drawn by combining pieces of data—a blind spot common among younger users as well [2, 63]—or were not certain how much inferencing currently occurs (*"If I were the evil genius, who had that record, I think I could [...] probably tell you more about yourself than you would know about yourself. Or I may be exaggerating, but not too much. [I: Do you believe anyone has the record on you?] I hope not, but, you know... I think most people would find it rather boring, but... [I: Do you think there's some evil genius exists somewhere in the world?] N-n-no, no. This is a hypothetical,"* P51).

Creation of detailed user profiles also enables secondary uses of the data [81], whether by the entity that collects the data or by someone the collector disseminates it to (see below). Our participants are aware of fraud, scams, and identity theft (25/46); targeted advertising (22/48); spam and telemarketing (17/46); and price and service discrimination (7/46). When we asked, "What does that device need to know about you to function properly?" participants' answers often jumped ahead from fulfilling functionality to secondary purposes. For example, when asked what her computer needed to know or collect, P77 responded, *"Cookies. It collects cookies. [I: Okay, what is that?] It tags certain sites that you go into, so the salespeople can send you the right kind of ads, mostly. That's what it says, that what cookies do.."*

In addition to financial fraud and run-of-the-mill online scams like phishing, 3 participants mentioned the potential for fraud on dating websites. This suggests that seniors' engagement with social media and online dating websites—often viewed as mostly relevant to younger generations—should be included in computer training programs for older adults.

Four participants mentioned that fears about information disclosure and/or re-identification limit their willingness to engage in online political discourse (*"I am always chatting about politics and, even on the phone, sometimes I hesitate because I know they cap all that information,"* P46; *"I would do a [Facebook] Like, or submit, and now I've decided not to do that because you just don't know what's being captured. But I really want to support those [political figures]. I don't think we know enough about what's being captured,"* P104).

Moreover, older adults are more engaged in health care activities than the general population [37], which increases their vulnerability to *medical* fraud and scams. Participants generally view medical staff as trustworthy recipients of sensitive personal information and described using online patient portals for managing and exchanging medical information. However, a few participants expressed concern that medical staff may misuse this data (e.g., to assign unnecessary or more expensive treatments, or for personal retaliation). Misuse can have severe consequences (*"I got a bill from the hospital for $26,000. They had padded it. [...] I can't prove that none of that stuff happened,"* P5).

Insecurity resulting from inadequate protections is another frequent concern. Twenty-four participants mentioned hacking, and six specifically mentioned viruses or malware.

**Information Dissemination.** With regard to information dissemination, older adults were primarily concerned with their personal information being sold for profit, or being disclosed with malicious intent to cause reputational damage, humiliation, or embarrassment.

Specifically, 11 participants discussed the possibility of information being sold and subsequently used for secondary or even malicious purposes (*"If it's confidential and private, I don't care if they have all my information. [...] As long as [...] it wouldn't be abused, or I'd get a bunch of salesmen calling me trying to sell a device or a pill or something,"* P10). Others' concerns were more general (*"I would just like to see some kind of safeguard [...] in the technology so that strangers [...] don't have access to knowing everything about you, because strangers don't really need to know,"* P47). Even if the initial intent is not malicious, disclosure of sensitive economic and health information can endanger benefits older adults might otherwise receive, such as social security, disability allowance, insurance coverage, and eligibility for senior housing or assisted living facilities.

Participants concerned about scams and fraud often recognized that the information being used by scammers (or even hackers) for illegitimate purposes may come originally from disclosures someone purposely made to legitimate recipients, demonstrating again the limits of users' control (*"I no doubt shared my social security number with some other benevolent entity [...] but that someone decided that that might be of value in the open market,"* P51).

Unlike with commercial data, in the few cases where participants mentioned specific cases of medical data having been shared in ways they saw as violating their privacy, it was usually obvious to them who had shared it and when.

However, participants also expressed the desire to balance

privacy and security with the benefits of data portability, especially with regards to healthcare, research accessibility, and legitimate access delegation (*"I wish [doctors] would share [my medical records with each other], but they don't. It's so compartmentalized that it's [...] really frustrating. [...] It's a benefit and it's a curse, [...] because [...] unless you tell them, [...] they don't know what is going on with the [other] doctors in your life,"* P46). For instance, the poorly defined legal role of informal caregivers generates annoyance about privacy and security protections and may erode privacy values (*"The privacy to me seems like overkill. The concern about the hoops I have to jump through to be able to order the wife's prescription or to speak for her. I know that there are lawsuit reasons [...] so they have to be so so so so careful. But I don't share that concern. It probably shows that I am naïve,"* P123).

**Privacy Invasion.** While the risks of physical attacks and reputation damage are not exclusive to the online world, participants noted that modern technology exacerbates them (*"When you are having a private discussion with someone, you ought to be able to feel that it's as private as those that are involved in it are willing to be, you know. You can't obviously be sure that they won't go blabbing it all to the next person they talk to, but, I wouldn't want technology doing that for me,"* P15). Participants were particularly concerned about location data and data about their in-home activities, which some saw as sources of compromising information that could facilitate physical attacks on them or their property.

A few of our participants were also concerned about interference in their decisions, such as the use of social media to interfere in the US elections (*"I think that they expected that Facebook information would be effective in addressing specific group of voters. When you think about it, it is not far-fetched. It is perfectly reasonable,"* P121).

### 5.1.2 Seniors' Views About Age-Based Differences

Some participants discussed beliefs about generational differences in privacy attitudes, or in privacy risks.

**Beliefs about Whether Seniors Are More Concerned about Privacy than Younger People.** We observed a dichotomy in seniors' views on age-based differences in privacy attitudes. Some participants (9/46) expressed fundamental beliefs about privacy. They explained that they grew up with the idea of privacy as a valuable human right, where information sharing has limits and rules defined by social norms—norms that some believe are changing across generations (*"...People say, 'Well, if you've got nothing to hide, why don't you tell them?' It's none of their business! [...] It's much less so in this new age: the millennials, they don't seem to be quite so concerned about it. But when I was growing up there was some very strong limitations on what you ask people, what you told people. [...] So it's a generational thing,"* P22).

Some other participants (4/46) expressed the contrasting belief that older adults do not need to worry about privacy as much as younger people do. Some ascribed particular reasons, such as not being concerned about job opportunities (*"If I was younger, it might hinder me from jobs or even benefits of some kind. But now I don't think it would inhibit me from benefits,"* P21).

So while some described changing needs or views over time (*"This may be a function of age because, at this stage of my life, I don't feel like I have great secrets or private information,"* P6), others view privacy as a constant (*"I'm old fashioned enough to know what privacy is and to value it. [...] If at my age I don't have a few things to hide from a few people, my life has been totally wasted,"* P113).

**Beliefs about Whether Seniors Are Seen as Attractive Targets.** Participants expressed some contradictory opinions about whether older adults are viewed as better targets for security and privacy attacks. Several participants believe older adults are specifically targeted because they are viewed (correctly or incorrectly) as vulnerable, easy targets, especially for social-engineering attacks. They attribute the targeting to assumptions about seniors' low technical literacy; lack of support (*"I think [the falsified bill] is because they think old people are stupid or they're not aware and I was there alone. I couldn't prove anything,"* P5); or gullibility (*"Because it's elderly are more fallible, or they're more trusting, so they take advantage,"* P7). At least one believes attackers make assumptions about their financial situation (*"Maybe he thinks I'm wealthy and [is] after my money,"* P13).

In contrast, a few older adults believe that attackers do not see them as "major consumers" (P110) and doubt that their information is useful enough to be exploited for commercial purposes (*"I think that I am not a focus of whatever these companies are looking for. They probably look at my data—if they look at it—and say, Oh, don't bother with her. She's too old to participate, or maybe doesn't have enough money, or I don't know what they think,"* P110).

### 5.1.3 Unrecognized Threats

Some older adults in our interviews did not purchase their own devices, and instead rely on used devices or public equipment and services. Few of those participants mentioned potential privacy and security threats associated with public or used devices, which we discuss below.

**Use of Public Devices and Services.** Older adults are less likely to own their own computers or smartphones than younger people [7], therefore, seniors are more likely to use public devices. Six participants mentioned that they use public computers (e.g., in libraries or at senior centers). Some use public medical devices; two participants mentioned that because they do not have blood pressure monitors at home, they *"go to Walgreens and other places, where they have free*

*checks. And I got it checked recently at a health fair,"* (P10). Privacy and security in such situations depends on what data is collected, how it is stored and used, and whether the devices and entities collecting it are subject to HIPAA [61].

Participants' use of public devices is usually motivated by either the high cost of purchasing a device or a lack of perceived utility in owning one, e.g., due to infrequent use. Infrequent use in turn amplifies security risks related to lack of skills and experience, e.g., in detecting malicious events or suspicious websites, links, or documents [71–73].

Few participants expressed concerns about public devices or public Wi-Fi networks, even though they are more likely to expose users to vulnerabilities such as malware infection, data leaks, and other privacy and security threats resulting from accidental shared access, shoulder surfing, and Wi-Fi spoofing Instead, most simply appreciated that someone else was maintaining the devices: *"That's another reason why I don't want a home computer. I go to the library, and if [the computers there] crash, they'll deal with it. [...] If I had one, and it crashed [...] I'd just leave it off. I don't want to have to pay for the repairs,"* P10. However, the effectiveness of maintenance is a function of the expertise and diligence of the person in charge and of the resources available at the public facility. Moreover, the security efforts of administrators can still be compromised by user behavior [9].

**Use of Second-Hand Devices.** Seven participants mentioned that they use second-hand devices given to them by family, friends, or neighbors (*"Grandpa gets the oldest phone. When they get upgraded, the phones trickle down. [...] I am thrilled with it, and it is too old for anyone else to use in that household,"* P121). The most common were smartphones, computers, tablets, and TVs, though one person mentioned a cleaning robot. Refurbished computers were also mentioned.

Reuse of such devices entails serious security and privacy risks, for both the previous owner (e.g., personal data disclosure, unauthorized access) and the new owner (e.g., malware and viruses). Moreover, access to technical support and security updates declines over time, further increasing vulnerability [70]. However, no older adults among our participants mentioned any potential risks from using second-hand devices, and only one mentioned that the previous owner reset the device, although it is not clear how effectively it was done (*"My friend did give me her old Mac. So I need to set that up. She wiped hers out. It's an older one, but she was using it for school, and she did video chats and everything on it, so it's very up-to-date. I don't need the latest,"* P36).

## 5.2 How Older Adults Manage Privacy and Security Risks

Similar to previous studies with older adults [16, 36, 47], our participants hold a range of attitudes about whether privacy and security concerns can be addressed in the current environment—which affects their attempts to mitigate those concerns. Some participants were pessimistic, believing that users have lost control over their personal information (*"I wish they would take the word privacy out of the dictionary. There is no such thing anymore. [...] I think it's the genie out of the box. I don't think it can be addressed,"* P43).

Such fatalism can result from a perceived lack of control and transparency, which leads to inertia against taking active security- and privacy-enhancing steps (*"I was thinking of cancelling my Facebook account but then I read that even if you're not a member, that they can get all kinds of information, so I don't know if I want to bother with that,"* P20). Another reason is a lack of confidence about having the knowledge and skills to protect one's own information (*"I'm not sophisticated when it comes to all these electronic gadgets and so I don't know what the possibilities are for control that is unavailable to hackers and thieves,"* P20).

Some participants explicitly attributed their attitudes to age (*"Don't forget, I'm old. And some things [...] you just sort of have to let go and you don't want to use your energy at it. [...] I want my information back and they say no, sometimes you just have to go ahead [...] Not everybody can fix everything. You just have to live with the consequences. That's why you shouldn't be saying nasty things on the Internet, because it comes back to haunt you and you can't fix them,"* P107).

Other participants are less fatalistic and discuss how privacy can or should be restored and protected (*"I value privacy. I don't necessarily want anyone who wants information about me to be able to get it too easily, and too cheaply. If they are going to get it, I want them to work for it, and pay for it, as a way of discouraging them,"* P113).

### 5.2.1 Passive and Active Mitigation Strategies

We categorize the end-user security and privacy management strategies participants talked about along a scale of *passive* to *active* approaches.

One of the most commonly mentioned (28/46) passive mitigation strategies is to limit the use of technology or to avoid it altogether—sometimes causing notable inconvenience to the non-user (*"When you get Uber, if you don't log out and sign off each time, they know where you are all the time. I don't like that, location. [...] [I: So, what are you doing about that, do you still use Uber?] No, I don't. [...] It is just that when I go [...] in the city, instead of getting on the bus it is easier call Uber and, you know, but I have discontinued that,"* P46).

Other passive strategies include using services and devices with good reputations or brand image, and just generally trying to be cautious. Relying on such passive strategies is a double-edged sword. For example, relying on caution is subject to overconfidence bias [1], and depends on the user's vigilance, knowledge, and skills in detecting malicious actions and predicting the consequences of their behavior [18]. At the same time, unfortunately, many participants mentioned simply

accepting or ignoring known risks.

Active mitigation strategies include configuring privacy and authentication settings, using protective software and services, and deleting or refusing to provide personal information. Many participants mentioned strategies that mitigated the consequences of violations rather than the causes, such as blocking unwanted contacts or content, or discontinuing their use of devices or services after experiencing privacy or security violations. We provide more details about these strategies, along with supporting quotes, in Appendix B.

#### 5.2.2 The Role of Usability and Learnability

Our participants often explicitly view themselves as vulnerable to privacy and security threats because they have trouble using and configuring new technologies by themselves and/or because they know less about how the technologies work.

**Usability, Learnability, and Risk.** Participants mentioned obstacles related to the usability and learnability of privacy and security functions. These obstacles often result from or are amplified by general usability issues.

Despite their prevalence, passwords suffer from well-known usability issues [66, 84, 93], such as needing to be memorized and changed (*"I have a list of [passwords], and sometimes the computer will remember them, which is helpful, and then sometimes not. I have it written down and sometimes they make you change the password and I forget to write it down,"* P6). Participants have a variety of strategies for dealing with this—including strategies that are commonly viewed as poor security practice (re-using, choosing simple/guessable passwords). Many participants have heard advice about good password practices, but cannot effectively implement all of that (sometimes conflicting) advice (*"I use the same password for everything and I have used the same password for years. Even though we have been advised not to do that. [...] It's hard enough for me to come up with a password that I can remember and not write down—they tell you not to write it down so I don't do that,"* P110).

In addition to authentication, participants mentioned potentially privacy-relevant usability issues like accidentally activating voice control on a phone, or not being able to figure out how to sync email to delete a message on all devices at once. In addition to a general feeling of having lost control or not having mechanisms to exert it, several participants doubt such mechanisms could ever be usable (*I: "What if the system will give you control over the information so you can decide who can access it? [...]" P: "That's just too much trouble. [...] By observing other people with computers, they are always messing up. [...] It's not just push a button and have it do what you want,"* P1).

**Delegation of Privacy and Security Management.** A related issue is that older adults often involve others in managing their privacy and security (e.g., configuring settings) [cf. 70]. They may even hand it over completely to family members, someone in their community, or technical experts (*"It's called Touch ID? [...] Yeah, I think I've heard of that, but my son did not set me up for that,"* P103). Delegation of security maintenance is a common practice among the general user population [29, 32], but due to especially limited digital literacy and experience, it may occur more frequently among senior users [12].

Older adults' need to turn to others for help with non-security-related technical issues (e.g., general setup and maintenance) can have security consequences. (Table 6 provides a general overview of older adults' tech troubleshooting strategies and issues that arise with each.) For example, sometimes older adults share account credentials with family members, friends, and (professional or volunteer) technical assistants [94]. One such community "technical assistant" commented: *"She didn't mind if I put [her] Amazon account in [my] phone, the credit cards and stuff, but I didn't want to get my Amazon account confused with hers, that's for sure,"* P123.

**The Consequences of Delegation for Learning.** Although relying on relatives and acquaintances to take care of technology setup and maintenance works for some participants, others discussed the difficulties such reliance can create. In particular, children or other family members might not have enough time to help, or when they do, might try to forestall further needs by discouraging older adults from fully using the technology. Limited explanations may leave older adults with an awareness of risks but few details on how they come about (*"My son is very good protect for my computer, not everybody can get it. It's very security for that. He just don't want me to check this, check that, get a virus. [I: So how does he protect...?] I don't know,"* P16). These issues emphasize the need for older adults to have independent channels for learning about and troubleshooting technology.

A few participants acknowledged explicitly that relying on others to set up and troubleshoot devices means they don't have much understanding about how they work (*"It's just part of my resistance to technology. [...] [The paid technician] is a smart guy and I don't have the patience to unravel it if it is not doing what it is supposed to do,"* P8).

A few said they just aren't interested in learning (*"I kind of just decided that I'm not interested in learning a lot of new technology,"* P77), but even those who are interested can find themselves falling back on asking others to solve problems for them (*"I belong to the computer club. [...] I've gone to their picnics a couple of times, but if you belonged to the club you have someone that will come and help you if you have problems with your computer. I don't have to know that much about it if I have a problem,"* P5).

#### 5.2.3 Sources of Information on Risks and Mitigation

Even participants who had not been targeted for specific privacy or security attacks seemed generally aware of potential

issues and described sources where they learned about risks.

News media are a common source [cf. 70]. Given the timing of the interviews (May–June 2018), Facebook's Cambridge Analytica scandal [13] came up frequently (*"Judging from the recent things that have come out with Facebook and Mark [Zuckerberg], I realize that whatever you type in, goes out"* P32). Several participants mentioned having heard about Alexa mistakenly sending a private conversation to a random contact in the owner's address book [46], as well as other stories about identity theft, data breaches, and data brokering.

Stories are sometimes accompanied by tips on how to avoid such scams or mitigate consequences of larger incidents, especially in publications for seniors such as the *AARP Bulletin* (*"Sometimes when [the service provider says], 'You should change your password. Your identity may have been stolen,' or something like that, then I would change my password. [...] Or, you know, on TV they would make that suggestion,"* P13). Data breach notifications from companies did not feature prominently in our interviews.

When the mitigation against a particular incident is fairly simple, these channels seem effective. However, more general or more complicated stories sometimes leave participants confused about the actual pathways data can take, and with a garbled or incomplete idea of how to protect themselves (*"Well I read in the paper that there are these search engines and they can get into computers [...] especially through Wi-Fi so I have Wi-Fi turned off,"* P108).

Another source of information about risks and mitigations is materials, classes, or lectures targeted specifically at older adults. Computer classes we saw advertised for seniors contained some privacy and security content. Generally, participants find computer classes beneficial (*"They give lessons, many, many classes every year on how to use your phone, or how to use computers, or how to use anything [...] and they're very good,"* P5). However, some noted that *"it's hard to know if [classes are] at the level that you need"* (P18), or find classes too difficult (*"I need like ABCs, 1-2-3s. It was not basic enough for me,"* P69).

Several participants mentioned having attended or heard about talks on how to avoid scams. For those classes, the relevance is generally clear (*"They have seminars on [...] how to avoid being scammed. [...] [I: Do you believe that it could happen to you too?] Yeah, why not, sure, but..."* P7). But in other cases, participants did not make the connection between lecture content and consequences for their data (*"Somebody came and talked about the cloud. What is it, what does it do, you know, that kind of thing. I went and I thought I don't need all this. [...] I just look things up and send a few emails and that's about it. I don't care about anything else,"* P5).

## 5.3 Notable Misconceptions and Blind Spots

We identified common misconceptions regarding technology, data collection and sharing, and protections that could lead to older adults' forming inaccurate privacy and security threat models, or increase their vulnerability to risks.

### 5.3.1 Uncertainty about Information Flows

Uncertainties about what data is collected, transferred, and used, and how, are common in the general population [8, 62, 90], and among our participants in particular. In addition to lack of transparency about data practices, lower technical awareness and experience can aggravate the proliferation of such misconceptions among the elderly.

As noted in §5.1.1, some participants expressed incorrect assumptions that technology only collects information users input themselves, or were uncertain about it (*"I like to think that the smartphone only has in it what I put in it. Now I could be dead wrong but I like to think that,"* P22; *"I don't see my phone capturing my data, unless—what I enter,"* P104).

In contrast, some assume that virtually everything is collected, shared, and retained, which can lead to fatalism or resignation (*"Apparently they can track, from cell phones and cell phone towers they have a record, they can piece together so much about you,"* P113 ).

In a couple of cases, misconceptions about data collection were due to uncertainty about which devices are Internet-enabled (*"I am assuming that [a smart speaker] is not really connected to the Internet. It has to do with information you put in, so I wouldn't worry about what information they had about me. [...] [I: It is connected to the Internet.] [...] Okay well I am wrong then, then it will know a lot more,"* P46). However, it was rarely so clear whether our respondents thought data collection and processing happen on-device or whether it is sent off-device. Although studies have shown [50] that this is an important distinction for users when asked about it explicitly, our study participants did not specify it unprompted.

Data flows in emerging technologies are especially opaque for older adults because they may be less familiar with the state-of-the-art sensors and algorithms, or with advances in artificial intelligence, than the younger population [80]. They may base their assumptions about how devices work—and therefore their privacy mitigations—on analogies with more familiar technology (*I: "What kind of information would you expect the devices to collect about you? [...] What about the smart speaker?" P: "Answering questions. I have begun to use this feature in the phone. [...] So, I guess what the smart speaker would do would be anything that the smart phone can do and then maybe more. I don't know what that might be,"* P60)—[cf. 69].

### 5.3.2 Uncertainty about Data Persistence

We also identified misconceptions about the effectiveness and extent of data deletion. A couple of participants said that when they delete a file or an email, they believe there is no longer any record of it, while in practice it is still locally stored and was simply moved to a Trash folder. The feedback they receive

from synced devices (when working correctly) reinforces this belief: when email is deleted on a computer, you can no longer see it on a mobile device, suggesting that it was deleted permanently (*"It is all connected. Once I delete it [on the computer], the phone is also,"* P7).

Several participants believe that data is overwritten, rather than stored permanently on the device or in a digital database. Sometimes these assumptions are based on analogies with older or more familiar technologies (*"I thought it was just [...] like recording over the tape [...] like where you used to tape programs from television. If you recorded over that tape, you wiped out pretty much what had been said or done,"* P35).

A couple of participants were also surprised about the duration of data retention (*"I hadn't even thought about [hearing aid apps] collecting [data], or where all that stuff goes. I think it's only me hearing it. Phew. Is a record of that around forever?"* P123).

Some participants assume that the information a device shows the user is a complete record of everything that device has collected (*"There's nothing that is recorded. [...] The only thing the phone would show is who called me,"* P110).

### 5.3.3 Blind Spots in Mitigation Strategies

Beyond data deletion, misconceptions about data flows and persistence, or about security mechanisms, may lead to older adults relying on other ineffective means of protection, or using protection strategies ineffectively.

Several participants mentioned not being sure about the effectiveness of their strategies (*"I gave money to a firm that said that they would provide some protection for my bank account, brokerage account. I don't know whether really that they would be that effective. [...] Probably a waste,"* P51). In extreme cases, the "security service" turned out to be a scam or ransomware attack (*"I got a call from some outfit that said that there was [...] some billing that had been done on my account from Russia. [...] And I said I didn't order that. [...] They persuaded me, which was an error on my part, to buy some service from them, and I bought the service and then I was told that that service offering was a scam,"* P20).

In contrast, some other participants may be overly confident about the effectiveness of the mitigation strategies they use, or due to lack of knowledge, consider less technologically advanced threat models. Such overconfidence may lead to neglecting security advice or reducing protection efforts: *"The nice thing about using Apple, is that there aren't hackers like there are with Windows. In Windows everything gets hacked so you have to have an anti-virus, an anti-something else, and you have to have the firewall. My Mac has two firewalls and that is all I need. [...] I think they come installed,"* P25).

Even when participants were aware of threats, they often did not know how to effectively protect against them. For example, P22 said *"I try to change my passwords regularly. And a lot of my passwords are so obscure I would be surprised if anybody could figure them out, although I know that they can be figured out. The references in my passwords are to things that nobody would associate with me. [...] So that's how I try and protect myself. I don't know how else to do it."* When choosing passwords, she does her best to try to make it harder for a lay person, presumably knowing some basic information about her, to guess it. However, such passwords may not be at all "obscure" for a hacker using brute force.

As we noted in §5.2.1, several participants mentioned strategies that mitigate privacy and security consequences, rather than the risks themselves. In some cases, they do not necessarily recognize that these strategies do not address the causes of the threat—or are not concerned that they do not. For example, a participant mentioned blocking telemarketing calls (*"I also have a call blocker on my phone. So I got rid of those unwanted calls [and] robocalls,"* P110). The participant was satisfied with the strategy, but of course a call blocker does not remove personal information from call lists.

A few participants acknowledged the ineffectiveness of mitigating consequences in addressing root causes, but said they felt helpless to find a better solution (*"You lose control once some outside agency has information. I am unable to stop the flood of phone calls whose origin and purpose I cannot imagine. The only thing I can do is what one daughter-in-law suggested—don't answer it,"* P69).

Unsubscribing, discontinuing, or simply abandoning a service can be as ineffective in addressing the root cause of the risk as mitigation of consequences. And when not done properly, it may even increase exposure. For example, abandoned accounts are often used for social engineering attacks and identity theft [88] (*"The other [incident of identity theft] almost had to be dishonest people that can view credit bureaus. Because a couple of accounts that we had zero balance on, we had cut up the credit cards, we had not closed the accounts,"* P123).

Finally, users only employ mitigation strategies when they have some awareness of the risks. Infrequently recognized risks are therefore infrequently protected against, for example, risks associated with public or hand-me-down devices (see §5.1.3).

### 5.3.4 Belief They Have Nothing to Hide

Echoing the "nothing to hide" fallacy [82], many participants feel that an honest person who has nothing to hide should not need to protect their privacy (*"I have no nefarious activities, so I have no problem,"* P121; *"I'm not that sensitive. I'm very 'open book' person,"* P31).

Similarly, some participants do not recognize the potential risks of data misuse (or underestimate its probability) if they do not view the information as sensitive or high-value (*"Who would really care how many steps a day I take? [...] I can't see how anybody could use that information to make money. [...] Unless maybe they wanted to sell me some exercise equipment, like a treadmill. [...] I don't see that as a realistic possibility of ever happening,"* P7).

One possible explanation for why these misconceptions occur is that participants often focus on the considerations of potential reputation damage and overlook broader security risks that could lead to material and financial consequences, or threats to physical safety. Although not unique to the older population [82], this misconception was quite common in our interviews, so we believe it is important to consider when designing privacy and security interventions for older adults.

## 6 Discussion and Implications

Below we summarize our findings, then use them as a basis for recommendations to providers of security awareness programs and education, and to technology designers. We also discuss potential future work.

### 6.1 Recap of Main Findings

Comparative findings in prior work show that distribution of privacy and security attitudes is similar across age groups [40, 41, 62, 91, 95], while privacy and security knowledge, behaviors, and risk levels differ [e.g., 34, 36, 41, 48, 58, 62, 73, 90]. Our results add depth to this picture, illustrating how certain privacy and security risks are amplified for older adults.

Amplification can be due to less knowledge and experience with technologies, decline in physical and mental abilities, and/or specific financial or living situations. For instance, we found that inhabitants of senior living facilities are particularly subject to surveillance, and often have to give up privacy and control of personal data. Our participants often reported using public and secondhand devices and public Internet access, yet they are not always aware of the potential threats involved. They are also concerned, confused, and often have misconceptions about data flows and risk mitigation strategies.

Participants provided insights on barriers to learning about, understanding, and using privacy and security protections, which are heightened by memory decline and physical limitations. In particular, we find that difficulty in using technology—whether older adults attribute it to user-unfriendliness or to their own lack of skill or knowledge—leads to a lack of self-efficacy about privacy and security. Therefore, addressing those barriers is an important basis for empowering older adults to use technology more safely and comfortably.

### 6.2 Suggestions for Awareness and Education Programs

We found that many older adults lack a nuanced understanding of newer technologies and the data they collect, leaving them especially vulnerable to privacy and security violations. Their particular concerns, misconceptions, and blind spots could be addressed through tailored training and educational efforts.

***Expand educational programming.*** Existing programming that older adults find valuable, such as computer classes, lecture series, or computer clubs, can be expanded. We recommend developing security and privacy materials *specifically designed for this age group*, in collaboration with trainers and older adults themselves. In addition to scams, such materials should address issues of most concern to older adults, such as surveillance, and misconceptions about data collection, persistence, and sharing. Engagement in social media, including dating websites, should not be overlooked. Potential risks of using public or hand-me-down devices, and how to mitigate them, should also be considered.

Targeted materials will allow those leading the classes to more easily tailor them to seniors' needs and knowledge—including making the necessary connections between technical facts and practical consequences, so that seniors better understand the relevance of the technical details.

***Leverage existing points of contact for outreach***. Privacy and security information for older adults can be disseminated via channels they already use to get help with computer problems (see Appendix C), as well as resources they look to for general help and advice, such as publications or websites directed at seniors [cf. 70].Vendors and computer-repair experts could make age-appropriate privacy and security "checkups" a standard part of setup or troubleshooting conversations with seniors.

### 6.3 Suggestions for Technology Developers

Participants often avoided or stopped using technology due to privacy and security concerns or violations, which also affect their intentions to purchase and use emerging technologies. Participants frequently linked their privacy and security behaviors to usability concerns. This finding is an important illustration of the direct economic incentive for technology designers, developers, and manufacturers to address the privacy and security concerns of older adults.

***Improve transparency and control, address misconceptions.*** Security and privacy controls should be designed to account for misconceptions common among older adults (see §5.3), to anticipate and address respective risks. Incorporating privacy controls *where the default is the most private setting*, as older adults rarely configure them [42], is a first, basic structural change.

Standardizing and being upfront about the types, amount, and granularity of information collected and shared may enhance older adults' awareness and reduce the likelihood they will discontinue use after being surprised by a perceived privacy violation. Device descriptions and apps should make clear when information is sent over the Internet (rather than processed on-device), and where possible should incorporate data-transmission indicators [38, 55, 100].

***Address usability issues and improve system design.*** Interfaces should be designed to optimize older users' ability to authenticate, configure settings, and accomplish other security tasks without errors in a reasonable time. For instance,

usability issues associated with aging-related ability declines, such as reduced vision and acuity, hand tremors, memory worsening, and lower skin conductance [17], may complicate authentication management [33] and may lead older adults to choose less secure mechanisms.

To address the identified usability issues, designers can rely on expansive knowledge and guidelines in that area [25]. For instance, they can add security indicators for "trustworthy" applications, or provide default configurations for data backup [21]. Designers and developers should focus on facilitating information management (e.g., editing and deleting personal records). Companies should involve older adults in the development process through participatory design and usability testing.

### 6.4 Future Work

Some of the patterns we identified in our exploratory qualitative study merit further systematic investigation, such as older adults' uncertainties about data deletion and retention, or their use of public and secondhand devices. Consequences of those behaviors could be assessed in controlled behavioral studies. In particular, it is not yet clear how the issues we identified affect older adults' privacy and security behavior as compared to the general population, or whether their security and privacy management strategies are more or less effective than those of the general population.

Older adults' use of emerging technologies, especially healthcare technologies, also warrants further exploration. While many of our participants used such technologies, or had heard of them, their use and knowledge was sufficiently heterogeneous that clear themes did not emerge. Further research is needed to examine specific privacy and security questions about older adults' use of these technologies in greater depth and at larger scale.

Finally, the measures we recommend should be tested "in the wild" to determine their efficacy. For example, we might test whether having targeted training materials for educational programs can positively impact older adults' privacy and security behaviors; or whether more transparency about data collection and sharing improves their comfort with using an app or device. Of particular importance would be age-specific usability tests of enhanced privacy and security controls, especially for new types of technologies such as healthcare and other monitoring devices.

## 7 Conclusions

As the population of older adults grows and turns their attention to technology, systems will need to be designed to enable informed choices, better control over personal data, and improved security for this user group.

Through semi-structured interviews with 46 older adults, we identified a variety of privacy and security attitudes and concerns, threat models and mitigation strategies, common misconceptions, and usability issues with currently deployed privacy and security controls. In general, the range of privacy and security attitudes, as well as the threat models and associated misconceptions mentioned by older adults in our interviews and reported in prior research, are also common among the younger population. However, our findings illustrate how older adults may be *particularly vulnerable* to certain risks and experience difficulties in mitigating them, due to age-related declines in abilities, and to their relative lack of technical knowledge and experience (shown in previous studies and confirmed here).

Emerging technologies featuring smart sensors or machine learning algorithms were especially concerning for our participants. Their data flows were difficult for participants to understand, likely because of their opacity. Participants specifically mentioned concerns about passive data collection (e.g., by smart speakers) and their privacy as bystanders (when other people's devices collect information about them).

Our participants often reported using public and secondhand devices and public Internet access, but were not always aware of associated privacy and security risks. They also mentioned concerns over the disclosure of sensitive financial and health conditions, which could be accelerated by the proliferation of e-health and health-monitoring systems. Participants mentioned concerns that such disclosures may endanger benefits they might otherwise receive, such as social security, disability allowance, insurance coverage, or eligibility for senior housing or assisted living facilities.

Residents of senior care facilities especially often acknowledged being resigned to the loss of privacy in exchange for care and safety. For seniors living independently, balancing the tradeoffs between care/safety and privacy is an open dilemma, as it conflicts with their desire for independence.

Finally, we found that one of the most commonly mentioned approaches to mitigating privacy and security risks was to avoid or limit using the technologies. This finding suggests that businesses offering devices or services targeted to or used by older adults may accrue economic benefits and gain a competitive advantage by considering the opinions and addressing the concerns of this population.

## Acknowledgments

We thank Joy Qiaoying Tang for recruitment help. CHI workshop participants provided helpful comments about the study, as did anonymous CHI and SOUPS reviewers. We also thank our participants, as well as the senior centers and care facilities that assisted in recruitment.

This work was supported by generous gifts from Cisco and Mozilla, by a grant from the Center for Long-Term Cybersecurity (CLTC) at U.C. Berkeley, by National Science Foundation grants CNS-1514211 and CNS-1528070, and by the National Security Agency's Science of Security program. Opinions, findings, and conclusions are those of the authors and do not necessarily reflect the views of the funders.

## References


[1] A. Acquisti, I. Adjerid, R. Balebako, L. Brandimarte, L. F. Cranor, S. Komanduri, P. G. Leon, N. Sadeh, F. Schaub, M. Sleeper, et al. Nudges for privacy and security: Understanding and assisting users' choices online. *ACM Computing Surveys (CSUR)*, 50(3):44, 2017.

[2] A. Acquisti and R. Gross. Predicting social security numbers from public data. *Proceedings of the National Academy of Sciences*, 106(27):10975–10980, 2009.

[3] Ageless Innovation LLC. Joy For All Companion Pets, 2018. https://joyforall.com.

[4] I. Altman. *The Environment and Social Behaviour: Privacy, Personal Space, Territory, and Crowding*. Brooks/Cole Publishing Company, 1975.

[5] K. B. Anderson. *Consumer fraud in the United States: An FTC survey*. Federal Trade Commission, 2004.

[6] M. Anderson. Smartphone, computer, or tablet? 36% of americans own all three. Technical report, Pew Research Center, 2015.

[7] M. Anderson and A. Perrin. Technology use among seniors. Technical report, Pew Research Center for Internet & Technology, Washington, DC, 2017.

[8] C. M. Angst and R. Agarwal. Adoption of electronic health records in the presence of privacy concerns: The elaboration likelihood model and individual persuasion. *MIS Quarterly*, 33(2):339–370, 2009.

[9] M. Arora, K. K. Sharma, and S. Chauhan. Cyber crime combating using KeyLog Detector tool. *International Journal of Recent Research Aspects*, 3(2):1–5, 2016.

[10] I. Azimi, A. M. Rahmani, P. Liljeberg, and H. Tenhunen. Internet of Things for remote elderly monitoring: A study from user-centered perspective. *Journal of Ambient Intelligence and Humanized Computing*, 8:273–289, 2016.

[11] S. Beach, R. Schulz, J. Downs, J. Matthews, B. Barron, and K. Seelman. Disability, age, and informational privacy attitudes in quality of life technology applications: Results from a national web survey. *ACM Transactions on Accessible Computing*, 2(1):5:1–5:21, 2009.

[12] V. Boothroyd. *Older Adults' Perceptions of Online Risk*. PhD thesis, Carleton University, 2014.

[13] C. Cadwaller and E. Graham-Harrison. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian*, 2018. Accessed on 19 September 2018.

[14] K. E. Caine, C. Y. Zimmerman, Z. Schall-Zimmerman, W. R. Hazlewood, L. J. Camp, K. H. Connelly, L. L. Huber, and K. Shankar. DigiSwitch: A device to allow older adults to monitor and direct the collection and transmission of health information collected at home. *Journal of Medical Systems*, 35:1181–1195, 2011.

[15] R. M. Califf and L. H. Muhlbaier. Health Insurance Portability and Accountability Act (HIPAA): Must there be a trade-off between privacy and quality of health care, or can we advance both? *Circulation*, 108(8):915–918, 2003.

[16] J. Camp and K. Connelly. Beyond consent: Privacy in ubiquitous computing (ubicomp). *Digital Privacy: Theory, Technologies, and Practices*, pages 327–343, 2008.

[17] N. Caprani, N. E. O'Connor, and C. Gurrin. Touch screens for the older user. In *Assistive technologies*. InTech, 2012.

[18] E. L. Carlson. Phishing for elderly victims: as the elderly migrate to the internet fraudulent schemes targeting them follow. *Elder Law Journal*, 14:423, 2006.

[19] B. Carpenter and S. Buday. Computer use among older adults in a naturally occurring retirement community. *Computers in Human Behavior*, 23(6):3012–3024, 2007.

[20] Cherry Home. https://cherryhome.ai, 2018. Accessed on 14 September 2018.

[21] E. Chin, A. P. Felt, V. Sekar, and D. Wagner. Measuring user confidence in smartphone security and privacy. In *Proceedings of the 8th Symposium on Usable Privacy and Security (SOUPS)*, page 1. ACM, 2012.

[22] J. Chung, G. Demiris, and H. Thompson. Ethical considerations regarding the use of smart home technologies for older adults: An integrative review. *Annual Review of Nursing Research*, 34:155–181, 2016.

[23] J. F. Coughlin, L. A. D'Ambrosio, B. Reimer, and M. R. Pratt. Older adult perceptions of smart home technologies: Implications for research, policy & market innovations in healthcare. In *29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1810–1815. IEEE, 2007.

[24] L. Coventry and P. Briggs. Mobile technology for older adults: Protector, motivator or threat? In J. Zhou and G. Salvendy, editors, *Human Aspects of IT for the Aged Population. Design for Aging*, pages 424–434, Cham, 2016. Springer International Publishing.

[25] S. J. Czaja, W. A. Rogers, A. D. Fisk, N. Charness, and J. Sharit. *Designing for Older Adults: Principles and Creative Human Factors Approaches*. CRC Press, 2009.

[26] G. Demiris, D. Oliver, G. Dickey, M. Skubic, and M. Rantz. Findings from a participatory evaluation of a smart home application for older adults. *Technology and Health Care: Official Journal of the European Society for Engineering and Medicine*, 16:111–8, 2008.

[27] M. Di Rosa, V. Stara, L. Rossi, F. Breuil, E. Reixach, J. G. Paredes, and S. Burkard. A wireless sensor insole to collect and analyse gait data in real environment: The WIISEL project. In B. Andò, P. Siciliano, V. Marletta, and A. Monteriù, editors, *Ambient Assisted Living: Italian Forum 2014*, pages 71–80. Springer International Publishing, Cham, 2015.

[28] K. Dobransky and E. Hargittai. Unrealized potential: Exploring the digital disability divide. *Poetics*,

58:18–28, 2016.

[29] P. Dourish, E. Grinter, J. Delgado De La Flor, and M. Joseph. Security in the wild: User strategies for managing security as an everyday, practical problem. *Personal and Ubiquitous Computing*, 8(6):391–401, 2004.

[30] A. Essén. The two facets of electronic care surveillance: An exploration of the views of older people who live with monitoring devices. *Social Science & Medicine*, 67(1):128–136, 2008.

[31] S. Fang, Y. Liang, and K. Chiu. Developing a mobile phone-based fall detection system on Android platform. In *2012 Computing, Communications and Applications Conference*, pages 143–146, 2012.

[32] A. Forget, S. Pearman, J. Thomas, A. Acquisti, N. Christin, L. F. Cranor, S. Egelman, M. Harbach, and R. Telang. Do or do not, there is no try: User engagement may not improve security outcomes. In *Proceedings of the 12th Symposium on Usable Privacy and Security (SOUPS 2016)*, pages 97–111, Denver, CO, 2016. USENIX.

[33] K. Fuglerud and O. Dale. Secure and inclusive authentication with a talking mobile one-time-password client. *IEEE Security & Privacy*, 9(2):27–34, 2011.

[34] V. Garg, L. J. Camp, K. Connelly, and L. Lorenzen-Huber. Risk communication design: Video vs. text. In S. Fischer-Hübner and M. Wright, editors, *Privacy Enhancing Technologies*, pages 279–298, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[35] V. Garg, L. J. Camp, L. Lorenzen-Huber, K. Shankar, and K. Connelly. Privacy concerns in assisted living technologies. *Annals of Telecommunications*, 69(1):75–88, 2014.

[36] V. Garg, L. Lorenzen-Huber, L. J. Camp, and K. Connelly. Risk communication design for older adults. *Gerontechnology*, 11(2):166–173, 2012.

[37] B. Gielen, A. Remacle, and R. Mertens. Patterns of health care use and expenditure during the last 6 months of life in Belgium: Differences between age categories in cancer and non-cancer patients. *Health Policy*, 97(1):53–61, 2010.

[38] S. Gray. Always on: Privacy implications of microphone-enabled devices. Technical report, Future of Privacy Forum, 2016.

[39] P. Gregor, A. F. Newell, and M. Zajicek. Designing for dynamic diversity: Interfaces for older people. In *Proceedings of the 5th International ACM Conference on Assistive Technologies*, pages 151–156. ACM, 2002.

[40] G. A. Grimes, M. G. Hough, E. Mazur, and M. L. Signorella. Older adults' knowledge of internet hazards. *Educational Gerontology*, 36(3):173–192, 2010.

[41] G. A. Grimes, M. G. Hough, and M. L. Signorella. Email end users and spam: Relations of gender and age group to attitudes and actions. *Computers in Human Behavior*, 23(1):318–332, 2007.

[42] R. Gross and A. Acquisti. Information revelation and privacy in online social networks. In *Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society*, pages 71–80. ACM, 2005.

[43] M. Haight, A. Quan-Haase, and B. A. Corbett. Revisiting the digital divide in Canada: The impact of demographic factors on access to the Internet, level of online activity, and social networking site usage. *Information, Communication & Society*, 17(4):503–519, 2014.

[44] E. Hargittai and K. Dobransky. Old dogs, new clicks: Digital inequality in skills and uses among older adults. *Canadian Journal of Communication*, 42(2), 2017.

[45] E. Hargittai and E. Litt. New strategies for employment? Internet skills and online privacy practices during people's job search. *IEEE Security & Privacy*, 11(3):38–45, May 2013.

[46] G. Horcher. Woman says her Amazon device recorded private conversation, sent it out to random contact. *KIRO News*, 2018. Accessed on 19 September 2018.

[47] D. Hornung, C. Müller, I. Shklovski, T. Jakobi, and V. Wulf. Navigating relationships and boundaries: Concerns around ICT-uptake for elderly people. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 7057–7069, 2017.

[48] M. G. Hough. Exploring elder consumers interactions with information technology. *Journal of Business & Economics Research (JBER)*, 2(6), 2004.

[49] L. L. Huber, K. Shankar, K. Caine, K. Connelly, L. J. Camp, B. A. Walker, and L. Borrero. How in-home technologies mediate caregiving relationships in later life. *International Journal of Human–Computer Interaction*, 29(7):441–455, 2013.

[50] I. Ion, N. Sachdeva, P. Kumaraguru, and S. Čapkun. Home is safer than the cloud! privacy concerns for consumer cloud storage. In *Proceedings of the Seventh Symposium on Usable Privacy and Security*, page 13. ACM, 2011.

[51] Jibo Inc. https://www.jibo.com, 2018. Accessed on 14 September 2018.

[52] S. Jones and S. Fox. Generations online in 2009. Technical report, Pew Internet & American Life Project, Washington, DC, January 2009.

[53] R. Kaye, E. Kokia, V. Shalev, D. Idar, and D. Chinitz. Barriers and success factors in health information technology: A practitioner's perspective. *Journal of Management & Marketing in Healthcare*, 3(2):163–175, 2010.

[54] B. Knowles and V. L. Hanson. Older adults' deployment of 'distrust'. *ACM Transactions on Computer-Human Interaction*, 25(4):21:1–21:25, 2018.

[55] O. Kohanteb, O. Tong, H. Yang, T. Saensuksopa, and S. Kazi. Guidelines for designing connected devices. Technical report, Carnegie Mellon University, 2015. Accessed on 26 February 2018.

[56] S. Kozina, M. Lustrek, and M. Gams. Dynamic signal segmentation for activity recognition. In *Proceedings of International Joint Conference on Artificial Intelligence*, volume 1622, page 1522, 2011.

[57] S. Ledbetter and L. Choi-Allum. Perspectives past, present, and future: Traditional and alternative financial practices of the 45+ community. Technical report, AARP, 2005. Accessed 2 May 2019.

[58] L. Lee, J. H. Lee, S. Egelman, and D. Wagner. Information disclosure concerns in the age of wearable computing. In *Proceedings of the NDSS Workshop on Usable Security (USEC '16)*. Internet Society, 2016.

[59] M. B. Lilly, A. Laporte, and P. C. Coyte. Labor market work and home care's unpaid caregivers: A systematic review of labor force participation rates, predictors of labor market withdrawal, and hours of work. *The Milbank Quarterly*, 85(4):641–690, 2007.

[60] U. Lindenberger, M. Lövdén, M. Schellenbach, S.-C. Li, and A. Krüger. Psychological principles of successful aging technologies: A mini-review. *Gerontology*, 54:59–68, 2008.

[61] D. D. Luxton, R. A. Kayl, and M. C. Mishkind. mHealth data security: The need for HIPAA-compliant standardization. *Telemedicine and e-Health*, 18(4):284–288, 2012.

[62] M. Madden and L. Rainie. Americans' attitudes about privacy, security, and surveillance. Technical report, Pew Research Center, May 2015. Accessed on 30 May 2019.

[63] A. McDonald and L. F. Cranor. Beliefs and behaviors: Internet users' understanding of behavioral advertising. Working paper. Accessed 3 May 2017: http://ssrn.com/abstract=1989092., 2010.

[64] A. McNeill, P. Briggs, J. Pywell, and L. Coventry. Functional privacy concerns of older adults about pervasive health-monitoring systems. In *Proceedings of the 10th International Conference on Pervasive Technologies Related to Assistive Environments*, pages 96–102, 2017.

[65] A. Melander-Wikman, Y. Fältholm, and G. Gard. Safety vs. privacy: Elderly persons' experiences of a mobile safety alarm. *Health & Social Care in the Community*, 16:337–46, 2008.

[66] W. Melicher, D. Kurilova, S. M. Segreti, P. Kalvani, R. Shay, B. Ur, L. Bauer, N. Christin, L. F. Cranor, and M. L. Mazurek. Usability and security of text passwords on mobile devices. In *Proceedings of the 2016 Conference on Human Factors in Computing Systems (CHI)*, pages 527–539, 2016.

[67] S. Mellone, C. Tacconi, L. Schwickert, J. Klenk, C. Becker, and L. Chiari. Smartphone-based solutions for fall detection and prevention: The FARSEEING approach. *Zeitschrift für Gerontologie und Geriatrie*, 45(8):722–727, 2012.

[68] H. Moghimi, J. L. Schaffer, and N. Wickramasinghe. Intelligent home risk-based monitoring solutions enable post acute care surveillance. In *Contemporary Consumer Health Informatics*, pages 399–412. Springer, 2016.

[69] A. Montanari, A. Mashhadi, A. Mathur, and F. Kawsar. Understanding the privacy design space for personal connected objects. In *Proceedings of the 30th International BCS Human Computer Interaction Conference: Fusion! (HCI '16)*, pages 18:1–18:13, Swindon, UK, 2016. BCS Learning & Development Ltd.

[70] J. Nicholson, L. Coventry, and P. Briggs. 'If it's important it will be a headline': Cybersecurity information seeking in older adults. In *Proceedings of the 2019 ACM Conference on Human Factors in Computing Systems (CHI '19)*, pages 349:1–349:11, 2019.

[71] G. Öğütçü, Ö. M. Testik, and O. Chouseinoglou. Analysis of personal information security behavior and awareness. *Computers & Security*, 56:83–93, 2016.

[72] D. Oliveira, H. Rocha, H. Yang, D. Ellis, S. Dommaraju, M. Muradoglu, D. Weir, A. Soliman, T. Lin, and N. Ebner. Dissecting spear phishing emails for older vs young adults: On the interplay of weapons of influence and life domains in predicting susceptibility to phishing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 6412–6424, 2017.

[73] Y. J. Park. Digital literacy and privacy behavior online. *Communication Research*, 40(2):215–236, 2013.

[74] PARO Robots U.S. Inc. http://www.parorobots.com, 2014. Accessed on 14 September 2018.

[75] S. T. Peek, K. G. Luijkx, M. D. Rijnaard, M. E. Nieboer, C. S. van der Voort, S. Aarts, J. van Hoof, H. J. Vrijhoef, and E. J. Wouters. Older adults' reasons for using technology while aging in place. *Gerontology*, 62:226–237, 2016.

[76] S. T. Peek, E. J. Wouters, J. van Hoof, K. G. Luijkx, H. R. Boeije, and H. J. Vrijhoef. Factors influencing acceptance of technology for aging in place: A systematic review. *International Journal of Medical Informatics*, 83(4):235–248, 2014.

[77] Qualcomm Technologies Inc. Home connectivity and integration, 2018. https://www.qualcomm.com/solutions/health-care/home-connectivity-and-integration. Accessed on 14 September 2018.

[78] A. Rao, F. Schaub, and N. Sadeh. What do they know about me? contents and concerns of online behavioral profiles. *arXiv preprint arXiv:1506.01675*, 2015.

[79] B. Reeder, E. Meyer, A. Lazar, S. Chaudhuri, H. Thompson, and G. Demiris. Framing the evidence for health smart homes and home-based consumer health technologies as a public health intervention for independent aging: A systematic review. *International Journal of Medical Informatics*, 82:565–579, 2013.

[80] K. Shankar, L. J. Camp, K. Connelly, and L. Huber.

Aging, privacy, and home-based computing: Developing a design framework. *IEEE Pervasive Computing*, 11(4):46–54, 2012.

[81] D. J. Solove. A taxonomy of privacy. *University of Pennsylvania Law Review*, 154:477, 2005.

[82] D. J. Solove. 'I've got nothing to hide' and other misunderstandings of privacy. *San Diego Law Review*, 44:745, 2007.

[83] StartUP Health. Digital health insights for the 50+ market: Prepared for the AARP. Technical report, AARP, 2014. Accessed 6 February 2019.

[84] E. Stobert and R. Biddle. The password life cycle: User behaviour in managing passwords. In *Proceedings of the 10th Symposium on Usable Privacy and Security (SOUPS 2014)*, pages 243–255, Menlo Park, CA, 2014. USENIX Association.

[85] J. Tao and H. Shuijing. The elderly and the big data: How older adults deal with digital privacy. In *2016 International Conference on Intelligent Transportation, Big Data & Smart City*, pages 285–288. IEEE, 2016.

[86] Theora Care. https://theoracare.com, 2018. Accessed 14 September 2018.

[87] L. Thomas, L. Little, P. Briggs, L. McInnes, E. Jones, and J. Nicholson. Location tracking: Views from the older adult population. *Age and Ageing*, 42(6):758–763, 2013.

[88] S. S. Tirumala, H. Sathu, and V. Naidu. Analysis and prevention of account hijacking based incidents in cloud environment. In *Proceedings of the 2015 International Conference on Information Technology (ICIT)*, pages 124–129. IEEE, 2015.

[89] D. Townsend, F. Knoefel, and R. Goubran. Privacy versus autonomy: A tradeoff model for smart home monitoring technologies. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4749–4752. IEEE, 2011.

[90] J. Turow, L. Feldman, and K. Meltzer. Open to exploitation: America's shoppers online and offline. Technical report, Annenberg Public Policy Center of the University of Pennsylvania, June 2005. Accessed on 3 June 2015.

[91] J. Turow, M. Hennessy, and N. Draper. The tradeoff fallacy: How marketers are misrepresenting American consumers and opening them up to exploitation. Technical report, Annenberg Public Policy Center of the University of Pennsylvania, June 2015. Accessed on 24 February 2018.

[92] J. Vines, S. Lindsay, G. W. Pritchard, M. Lie, D. Greathead, P. Olivier, and K. Brittain. Making family care work: Dependence, privacy and remote home monitoring telecare systems. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '13, pages 607–616, 2013.

[93] R. Wash, E. Rader, R. Berman, and Z. Wellmer. Understanding password choices: How frequently entered passwords are re-used across websites. In *Proceedings of the 12th Symposium on Usable Privacy and Security (SOUPS 2016)*, pages 175–188, Denver, CO, 2016. USENIX.

[94] M. Whitty, J. Doodson, S. Creese, and D. Hodges. Individual differences in cyber security behaviors: An examination of who is sharing passwords. *Cyberpsychology, Behavior, and Social Networking*, 18(1):3–7, 2015.

[95] W. Wilkowska and M. Ziefle. Perception of privacy and security for acceptance of e-health technologies: Exploratory analysis for diverse user groups. In *Proceedings of the 2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*, pages 593–600. IEEE, 2011.

[96] S. L. Willis, K. W. Schaie, and M. Martin. Cognitive plasticity. In *Handbook of Theories of Aging*, pages 295–322. Springer, 2009.

[97] World Health Organization. World health statistics. Technical report, World Health Organization, 2014. Accessed on 4 September 2017.

[98] X. Yu. Approaches and principles of fall detection for elderly and patient. In *Proceedings of HealthCom 2008: 10th International Conference on e-Health Networking, Applications, and Services*, pages 42–47, 2008.

[99] E.-M. Zeissig, C. Lidynia, L. Vervier, A. Gadeib, and M. Ziefle. Online privacy perceptions of older adults. In *International Conference on Human Aspects of IT for the Aged Population*, pages 181–200. Springer, 2017.

[100] E. Zeng, S. Mare, and F. Roesner. End user security and privacy concerns with smart homes. In *Proceedings of the 13th Symposium on Usable Privacy and Security (SOUPS 2017)*, pages 65–80, Santa Clara, CA, 2017. USENIX.

[101] K. Zickuhr and M. Madden. Older adults and Internet use. Technical report, Pew Internet & American Life Project, June 2012.

# A Privacy and Security Risks and Concerns

Table 3: Privacy and security risks and concerns based on Solove's taxonomy [81].

| Group of risks | Examples | Supporting quotes |
|---|---|---|
| Information collection | Tracking of online browsing. | *"The computer with cookies, they know where I go. They have data about me that I really don't like them having. This whole idea of computers knowing how the users are using the computer and gathering that data and then selling that data to others who make money from that data. I have real difficulty with that,"* P60; *" I know that there are a lot of people are watching what you do on the computer so I don't do anything,"* P5. |
| | Video and audio monitoring; data collection by wearable and context-aware sensors; surveillance including personal stalking, broader government and political surveillance, and monitoring of older adults by family members, medical care staff, or senior facility management; passive audio and video collection by phones, computers, fall detectors, smart TVs, voice assistants, and home-control systems. | *"We Jews don't face the repression in this country today that we faced in my parent's generation, okay? [...] [But] I am never completely far removed from thoughts of political repression. That's why I talk about surveillance,"* P113; *"These Alexa things [...] I guess it's always on, and always capturing [my data],"* P104; *"A person has some kind of a [...] voice assistant and that [...] record his private conversation and send it to somebody else. So I don't think it is a safe thing to have. I would throw it out of the window,"* P37; *"With the new smart televisions if you know, like with the computer too, they have the camera that they can look at you. [...] Some people cover up the camera with a piece of paper or tape. I am not quite that paranoid,"* P33. |
| | Violation of bystanders' privacy, especially by voice-activated, video-monitoring, and other context-aware systems. | *"I guess it's like an invasion of privacy. [...] When someone puts you in a room, they should tell you that there's a recorder there,"* P37. |
| Information processing | Data aggregation; individual profiling; targeted advertising. | *"They know everything you are doing, they know what you are looking at, they know what you are, you know, searching for and everything else. [...] One thing if you are looking at it on the computer, but then if you are talking to somebody and you make a remark about somebody or something or about politicians or something, well somebody could actually gather all that data and use it and say, well this person is a nasty democrat or left-wing or right-wing or whatever, so that is the only thing concerning, about the smart speaker especially,"* P33; *"Everything you buy, everything you look at, even, you know, if I go on Amazon and I look at something, then I'll see an ad for it on Facebook. [...] I don't like all these ads,"* P108. |
| | Telemarketing, spam (e)mail and calls, and other unsolicited marketing. | *"Oh yeah, you get a lot of weird calls when you are a senior in a rest home,"* P108; *"When you go on to these other sites looking for something then you get a barrage of emails afterwards. And I either delete them and if they keep on coming I try to find the place I can unsubscribe to them,"* P110. |
| | Fraud and scams (including medical contexts); phishing; identity theft by phone, email, and through social media (including dating websites). | *"They could probably scare me. They could say you have cancer, or you have something that we can't cure, or you need a surgery that you don't need. [...] Just for profit. [...] Let's say that they are a doctor who doesn't accept Medicare or your [insurance] plan, and they say well you have to pay for this out-of-pocket because you think you have cancer and you need a special medication or something,"* P46; *"Somebody was using [my friend's] Kaiser[3] number and getting services at another Kaiser location, and then she started getting these weird co-pay bills and discovered [her medical identity was stolen],"* P71. |

Continued on next page

[3] Kaiser Permanente is a major U.S. health care and insurance provider.

**Table 3 – continued from previous page**

| Group of risks | Examples | Supporting quotes |
|---|---|---|
| | Unauthorized access to personal information, e.g., by hacking, accidental shared access, or abuse of power. | *"People that shouldn't have access to your records who are in an official capacity could, you know, use information about you that they happen to see. [Say] somebody works at the DMV*[4] *and they looked up address of ex-girlfriend [...] and then they've got out and hurt that person,"* P71. |
| | Price and service discrimination; jeopardizing benefits older adults might otherwise receive, such as social subsidies, disability allowance, insurance coverage, or eligibility for senior housing. | *"[My personal information] might be used to influence my insurance company to raise my rates,"* P22. *"I am grandfathered in. [...][The director of the senior residence] would like to get us out. She's attempted in the past. [...] We have to [...] report income every year. [...] And when she first saw mine, she was very uppity about why the hell I was there. [...] But if I paid current rent [...], I'd be homeless in 10 years. And she said, 'Well then you would qualify for here,"'* P36. |
| | Viruses; malware; ransomware. | *"You just can't tell what's a virus and what's authentic. It does make me, I got a virus on my computer from something and got scolded. For falling for something [...] both [by] my son and the repairman,"* P18. |
| | Data integrity; mistakes and errors in personal records. | *"You wouldn't want somebody putting misinformation in your record. Or [...] changing information in there,"* P71. |
| Information dissemination | Disclosure; data breaches; selling of personal information to third parties. | *"It's mostly other companies that I never, I really never shopped in the first place that send me emails. [...] Those are the ones that I always want to get rid of. [I: And how do you think they got your information, then?] I'm sure it was shared by others. In fact, I know for sure that the [state] Department of Motor Vehicles sells your name and address. And I don't know what else they sell,"* P110. |
| Privacy invasion | General concerns about violation of privacy as a fundamental human right; interference in personal decisions. | *"If other people can find out things about you that you don't tell them yourself, yes, I would consider that intrusive,"* P1; *"I have personal knowledge about this type of situation in a family where somebody wants to [...] try to make a case that somebody is incompetent and the only way for them to do that would be, you know, to provide some sort of proof,"*, P47. |

Table 4: Consequences of privacy and security violations.

| Consequences | Description | Supporting quotes |
|---|---|---|
| Financial and material losses | Material and financial losses, including robbery or property damage. | *"Will they get something from my pattern? Would they track my daily activities? [...] So they can break into my house. I'm worried about that,"* P103. |
| Threats to health or physical security. | Health impairment, injuries, and threats to life or safety. | *I: "How do you think this recorded conversation or medical records or location or activity level or anything can be misused?" "P: Well people can spy on it and then they want to come in and kill you. They want to know when there is no sound and you are asleep, then they come in,"* P37. |
| Intangible consequences | Emotional, social, or ethical consequences, such as reputation damage, formation of stigma, social judgment, or anxiety. | *"[They could say] 'Oh he has a smart phone and he's [...] going to a meet up place where guys meet up.' [...] It could be interpreted. Surmised [that] I'm [a] bisexual guy. [...] I don't know exactly how they would take it. Or getting rebuffed and stigmatized,"* P9. |

[4]The Department of Motor Vehicles (DMV) is a state-level government agency that administers vehicle registration and driver licensing.

## B Mitigation and Management Strategies

Table 5: Mitigation and coping strategies.

| Passive strategies | Description | Supporting quotes |
|---|---|---|
| Limiting or avoiding technology use | Not keeping personally controllable data online or in digital format; not engaging in activities like online banking, online shopping, or social media; not using devices in general. | *"I guess whatever [a computer] knows about me is whatever I have put in or somebody else has. [...] That's why I continue to not use online banking or online payment services,"* P25; *"I don't want [my financial information] on the Microsoft cloud, I don't want it on the Apple Cloud. I want it on a hard drive that I know is on that computer and the portable hard drive that is hooked up. I don't use a wireless backup, a cloud back up,"* P123. *"I am not counting on protection of my privacy. [...] I do not use Facebook, I do not use any social media at all,"* P121. |
| Using services and devices with good reputations or brand image | Reliance on manufacturers to ensure security protection; confidence that a product with reputable name is safe against security threats. | *"I trust Apple more than most anyone. [...] If you sign into iCloud, if you have that two-layer security turned on, whatever that is called, that's pretty secure stuff,"* P123; *"The nice thing about using Apple, Linux is the system I use, is that there aren't hackers like there are with Windows. Windows everything gets hacked so you have to have an anti-virus, an anti-something-else, and you have to have the firewall. My Mac has two firewalls and that is all I need,"* P25. |
| Trying to be cautious | Self-censoring transmitted content. | *"I'm aware that there is no privacy, so I would never say anything on my phone or put anything in an email that I felt was in some way exposing me to liability or whatever,"* P121; *"I would do a [Facebook] like [for political figures], or submit [...] and now I've decided not to do that because you just don't know what's being captured. [...] And not like anything bad's going to happen to me, you know what I mean? [Not like] I'll get stopped at the border or something.,"* P104. |
| | Developing and applying methods to recognize suspicious content or untrustworthy intentions, e.g., in online dating. | *"I try to be very careful with what I get on my email. I don't indiscriminately open every message I get. If it's not a name I recognize, I delete it, I don't even open it,"* P110; *"He's real rich, and he's so handsome. [...] He writes down pages and pages, [...] as far as 'You make my life complete' and he hasn't met me yet! [...] So after a few times, I said, 'You're too good to be true,' and that sets off a red flag,"* P13. |
| Accepting or ignoring risks | Viewing personal information as an unavoidable trade-off in exchange for safety, or "free" Internet services; avoiding the high financial cost, time and effort, or questionable effectiveness of a remedy. | *"One of the advantages of living in [an assisted living facility] is that they have your complete records, and are in touch with your doctor,"* P121; *"Facebook is free. In exchange [...] you give up all this information because it goes to advertisers. [...] So lots of different things that used to be [...] technically free, they never really were, they were all monetized,"* P71; *"If you give to one pet organization they probably pass your name along to others. You know. I just have come to expect that. That's a part of the electronic age,"* P110; *"Some things you have no control over and can't do anything about. And also some things that you shouldn't be spending your time to do. [...] If you can't fix it or get them to fix it, or don't do anything about that, 'I want my information back' and they say no,"* P107. |
| **Active strategies** | | |
| Using or enhancing authentication mechanisms | Using passwords as required; screen-locking PINs; two-step verification; biometric authentication. | *I: "How do you keep track of your password? [...]" P: "I have this file for every company, everything that I use a password, I have it down there. [...] But I try to change them once in a while,"* P13. *"I think if you have the special connect with the hospital or the clinic and you have the special, you have the PINs or the security code, I think it's okay because the other picture over there, you can see which doctor you want to talk to,"* P16. |
| Configuring settings | Refusing location sharing permissions; deleting cookies; managing audiences. | *"I only have GPS on my phone when I need it. Nobody needs to know where I am—like MoviePass. MoviePass.com apparently wants to know where you are"*, P104, *"I have set [Mozilla] Foxfire [sic] so that when I close [it], all the cookies are deleted,"* P108; *"You can have a universal setting [on Facebook] and then when you post you can change that for the particular post,"* P108. |

Continued on next page

**Table 5 – continued from previous page**

| Active strategies | Description | Supporting quotes |
|---|---|---|
| Protective software and services | Anti-virus; ad-blocking and anti-tracking programs. | *"Well after being hacked, I don't know if [...] it can really be secure. I mean you purchase this anti-virus stuff that you put on there but it seems like they are not able to do the work. If someone is bent on wanting to get into your data or whatever device. That is pretty freaky,"* P5. |
| Active management of personal information | Refusing to provide personal information; providing fake information or dummy email addresses; deleting personal records. | *"I never give them my correct personal information. Just email. And a email is just set up for [contests],"* P104; *"As I learn how to use it, I will delete what I didn't feel comfortable with. If it wasn't applicable to me. [...] If it wasn't useful information, I would delete it,"* P60. |
| Discontinuing services | Unsubscribing, discontinuing, or simply abandoning a problematic service. | *"If you put the freeze [on your account with a credit bureau], nobody can use your name to apply for new credit card. And then if you know something happens, just close the account, right?"*, P103; *"My daughter got me a Facebook account. [...] When she set it up, we went on it together, and I haven't been back,"* P15. |

# C Troubleshooting: Who Older Adults Turn To

Table 6: Troubleshooting resources used by participants.

| Troubleshooting resources | Comments | Quotes |
|---|---|---|
| Providers | Older adults in our study most frequently look for help from the service provider, the device manufacturer, or the store/vendor. In some cases, they find these sources satisfactory. | *"The iPad, I went down to Apple, they're always crowded but I went very late and um, I was there for like an hour and a half and they got it—you know, they updated it. So, I think they do a good job because as you say, if you buy equipment and you can't get it to work, it's very frustrating,"* P44. |
| | However, some expressed reservations about how much time it could take to get help, or irritation at having to deal with chat bots or non-native English speakers. | *"What happens frequently [...] you have a question, an issue, and you're offered live chat. [...] Which really isn't a chat, it's sort of a messaging. I hate it. I cannot, I won't go near it. [...] I want to deal with humans,"* P15. |
| Personal network | The first call many participants make is to children, relatives, neighbors, or others in their personal network. Some of these helpers are (or were) computer or IT professionals; in other cases, they may only just know more than the participants themselves. | *"I have a guru that lives in southern California. I mail him stuff, we just sent him my computer, the hard drive just died. [This guru] it's my son! He's my computer expert. I want a new computer. I have a new computer. He sends it up, all installed. All I have to do is plug it in,"* P77. |
| Freelance or volunteer technicians | Several participants also mentioned computer experts they frequently call on—either paid technicians, or volunteers at a senior center or library. Some volunteers are also older adults, who provide help to others in their senior programs or housing facilities. | *"Okay, depending on how bad a technical issue it was, we used to have a guy that, our place provided somebody that used to come to help people with technology. You know, or to teach them how to get around,"* P36. |
| Do it themselves | Participants may first try to set up the device or solve the problem themselves, either relying on their prior knowledge or searching online for how-to videos, instructions, or help forum postings. | *"I figure them out [the technical issues]. If I don't figure them out, there are one or more persons that I could call,"* P21. |
| | Less frequently, they may try to find answers in the instruction manual, but some find manuals confusing or opaque. | *"The instructions have to be a, b, c, d, and e. You can't just do a and b and skip c and go to d and e. [...] Smartphones don't always tell you everything that the phone can do. You have to figure it out yourself. I have trouble with that only because it's so complex,"* P35. |

# Evaluating Users' Perceptions about a System's Privacy: Differentiating Social and Institutional Aspects

Oshrat Ayalon, *Tel Aviv University* Eran Toch, *Tel Aviv University*

## Abstract

System design has a crucial effect on users' privacy, but privacy-by-design processes in organizations rarely involve end-users. To bridge this gap, we investigate how User-Centered Design (UCD) concepts can be used to test how users perceive their privacy in system designs. We describe a series of three online experiments, with 1,313 participants overall, in which we attempt to develop and validate the reliability of a scale for Users' Perceived Systems' Privacy (UPSP). We found that users' privacy perceptions of information systems consist of three distinctive aspects: institutional, social and risk. We combined our scale with A/B testing methodology to compare different privacy design variants for given background scenarios. Our results show that the methodology and the scale are mostly applicable for evaluating the social aspects of privacy designs.

## 1. Introduction

System designs that do not meet the users' privacy expectations can startle users and lead them to abandon the system altogether [16, 20, 41, 50]. For example, in Felt et al. study, a participant reported about uninstalling an app after it had used his/her contact list information to send spam texts and emails [20]. These examples of mis-design highlight the importance of designing systems with privacy from the ground up, as promised by the Privacy-by-Design (PbD) approach. It calls for implementing privacy mechanisms in the systems at the initial stages of the development process to create privacy-respectful systems in advance [13, 38]. While PbD is part of official guidelines by the FTC and by the recent European General Data Protection Regulation (GDPR) [21], it is also criticized of being too focused on compliance to privacy regulation, rather than on providing the best privacy design to the users [66]. As a response, Koops et al. argue for broadening the envelope of PbD, fostering "the right set of mindset of those responsible for developing and running data processing systems." [34]



*USENIX Symposium on Usable Privacy and Security (SOUPS) 2019.* August 11 -- 13, 2019, Santa Clara, CA, USA.

End-users' long-term concerns and expectations are not always considered in the process of designing the privacy characteristics and features in systems. Therefore, we argue that privacy-by-design processes should take a more user-centered approach, and should put a stronger emphasis on involving users' views and feedback. Studies have shown that developers consult other developers [7, 23] or with the Chief Privacy Offices (CPOs) [7, 8] when designing for privacy. However, keeping design in narrow "professional" circles is highly problematic. As danah boyd argues, it is crucial to understand the social and cultural factors involved in the context of the way systems are used: "technologists assume the most optimal solution is the best one, but this tends to ignore a whole bunch of social rituals that have value." [10].

Leaning only on internal testing before launching a new system or feature can end up in systems that mismatch users' privacy expectations. This is particularly important as end-users' privacy expectations are not only about the way their data is handled between them and the system (an aspect known as institutional privacy [51, 52]), but rather, expectations also relate to social privacy: how systems allow managing relationships between end-users, and the complexity that sharing and hiding information plays in these relationships [35, 51, 52]. Legal frameworks hardly address social privacy, as long as users have agreed to the terms of service [9]. However, consent does not necessarily mean that users' expectations are met, as can be evident in previous privacy designs that included consent but surprised users [20, 50].

To effectively receive feedback from end-users about their perceived privacy of the system, there is a need to reliably measure their observations. Many works have suggested methods and scales to measure people's privacy attitudes and concerns [15, 25, 28, 42, 43, 56, 60, 68]. Some of these studies have focused on systems' privacy evaluation. For example, Suh et al. have created a scale that measures users' burden in computing systems, which includes a specific construct to evaluate the system's privacy [60]. However, these studies have mainly dealt with institutional privacy, rather than social privacy [29, 30]. Our study extends this strand of research by working towards a tool that is built to measure how users perceive a particular design. Currently, there is no generic scale that can point to a feature that is considered as alarming and inappropriate by the end-users in any given system design.

In this study we attempted to develop and validate the reliability of a novel privacy scale that adds a social aspect which highlights the information flow between people using the system. We used the scale to explore whether the usage of A/B testing, also known as a controlled experiment, is applicable for privacy evaluation purposes.

We conducted a study with two major stages: 1) seeking to develop users' perceived privacy scale, and 2) comparing privacy designs by using the scale. We began with the scale development, in which we recruited 459 participants via Amazon Mechanical Turk (AMT). To validate the scale we used several methods including principal component analysis (PCA), exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). At the second stage of the study, we used the scale to compare two different privacy designs of given background scenarios, borrowing the controlled experiment methodology. We recruited 858 participants via AMT and found significant differences between the designs in three out of the five background scenarios. The study results show that a controlled experiment can be extended to privacy evaluation, mostly for social privacy. In the same manner they show that our scale is partially sensitive enough to differentiate between the two design variants, according to the differences in social and institutional information management and controls, as well as the overall risk users feel involved in using the system.

## 2. Background

### 2.1. Privacy by Design

Privacy, as defined by the sociologist Alan Westin, is "the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others." [63]. In contemporary information systems, the principle of control manifests itself in several ways. Specifically in the context of online social networks' (OSNs), several studies have found that there are two ways in which users think about privacy and their control over their data: 'institutional privacy', which reflects the data relationship between users and the system, and 'social privacy', which reflects the data relationship between users mediated by the system [35, 51, 52]. Raynes-Goldie has found that OSNs users are more concerned about controlling access to their information by other people, rather than about how companies are using the data [52].

Privacy-by-Design is an approach that advocates mitigating privacy threats from the very beginning of the system development, rather than by adding privacy-enhancing technologies after the fact [13, 38]. Recently, PbD is becoming a central tool in regulatory frameworks, endorsed by the U.S. FTC and the new European GDPR. The latter requires data controllers to implement "data protection by design and by default." [21].

Several studies have shown that when designing for privacy, developers mostly focus on security and protection against external entities [6, 23]. In the cases of encountering privacy issues, developers often see these as someone else's responsibility [23] or seek advice from other developers or legal and managerial entities inside their organizations [7]. Wong and Mulligan call to "bring design to the PbD table," enriching PbD practices and research with design as a conception of a process, rather than as a mere tool for implementing objectives [65]. We add to their call and focus on the users' aspect. We argue that PbD processes consistently neglect the end-user's perspective that should be considered during the development process along with efforts of compliance. Only a handful of PbD white-papers have recommended involving users or receiving feedback from them (the suggestion to use focus groups by the UK Information Commissioner's Office [27] is an exceptional example). Involving users is never a mandatory requirement in formal PbD processes, and there are no proven methods to carry out user feedback in scale.

### 2.2. Controlled Experiments in User Experience

Controlled experiment for evaluating user experience, popularly known as A/B testing, is a methodology that is used to have a better understanding of the advantages and disadvantages of different designs. In a controlled experiment, users are randomly exposed to one variant (for example, two different shapes of a button), in a persistent context, in which the difference between the variants is minimized. Referring to the previous example, the variants will differ only or mostly in the tested button [32]. Controlled experiments are essential tools for web-facing companies, using such experiments to gain valuable customers feedback in a short time. The experiments purposes are varying, including mostly monetization, and testing usability improvements [31]. In our study, we used a controlled experiment to compare privacy designs alternatives. We want to explore the applicability of using this methodology for privacy evaluation purposes, also noting that there are different privacy aspects, as social and institutional, which might behave differently.

### 2.3. Measuring Privacy Attitudes

Several scales that consider different aspects of privacy were developed over the years. However, these scales were mostly developed to measure individuals' privacy attitudes, as their personally privacy concerns, rather than to evaluate systems. One of the most used scales was developed by Malhotra et al., who developed the Internet Users' Information Privacy Concerns (IUIPC) scale [42]. Dinev et al. took a different perspective in which they referred to privacy as a state within a certain context. Therefore, rather than

measuring privacy concerns, they measured perceived (state of) privacy [15].

Previously mentioned scales measured users' privacy approaches, while not measuring privacy perceptions in the context of a specific system. Other studies developed scales that included constructs that measured system's privacy [28–30, 43, 60, 68]. However, these studies have focused on institutional privacy aspects, lacking the perspective of social privacy. Considering the substantial participation of users in social media, Page et al. called privacy researchers to refer to social privacy concerns [49]. Several studies in the networked privacy domain developed privacy scales that were specific to privacy in the context of social networks. Stutzman developed an instrument to elicit the users' attitudes about the access to personal information by different people with different relationships [59]. Young and Quan-Haase used their privacy protection strategies instrument [69] to find that Facebook users' have developed privacy protection strategies and that they are mostly used to protect users against social privacy threats and not against institutional privacy [69]. Considerable research was dedicated to understanding privacy in social networks [36, 44, 61]. However, as social privacy is a concern in every collaborative system, there is a need to understand user expectations regarding the way a given system allows users to manage information sharing and privacy between users.

#### 2.4. Research Objectives

Taking the approach of Suh et al.s, in creating a scale for measuring user burden in systems [60], we aim to fill a gap in measuring the users' perceived privacy of a tested system. Moreover, it is unclear how various aspects of the system's privacy, including social and institutional privacies, affect users' perceived attitudes towards the system.

In this study, we aim to understand whether it is appropriate to use a controlled experiment, user-centered design methodology, to evaluate privacy design.

The first step was to define a reliable measurement, with which we can quantify the system's privacy, as it perceived by the users. Our literature review had brought us to investigate two distinct privacy aspects:

**H1.** Users' perceived privacy of a given system consists of two distinct aspects: social privacy and institutional privacy.

Once we have a reliable measure, we can explore whether controlled experiment methodology is applicable to compare privacy designs:

**RQ.1.** Can controlled experiment methodology differentiate between privacy designs?

**RQ.2.** Does the controlled experiment methodology applicability depend on the tested privacy aspect (social or institutional)?

## 3. Initial Scale Design

The goal of our scale is to measure end-users' perceived privacy of a specific information system, as we named it: Users' Perceived Systems' Privacy (UPSP) scale. We strongly based our scale on previous studies that created privacy scales. Some of the studies presented general privacy scales [4, 15, 19, 25, 42, 56] and others were specified to privacy in the context of OSNs [36, 44, 59, 61, 69]. Based on the literature review we identified a gap of a missing scale to measure perceived privacy from a social aspect, and that is aimed to evaluate an information system. Therefore, we attempted to create a scale that covers simultaneously both institutional-related aspect, which refers to privacy aspects between the user and the system, and social-related aspect, which refers to privacy aspects between the user and other people.

At the first stage of the scale development, we created a list of questions to represent institutional-related aspect. We chose several constructs that appeared on the previous general (i.e., not OSNs specified) privacy questionnaires and made adaptations when required, to represent questions about users' perceived privacy of the system. The chosen constructs were: perceived information control, confidentiality, importance of information transparency, secondary usage, data deletion, perceived privacy risk, and information sensitivity.

At the second stage, we developed new social-related questions based on the previously mentioned constructs and based on OSNs' specified privacy questionnaires. The social-related questions included two additional constructs, according to the original study upon which the questions are based on protection strategies [69] and identity sharing [59]. See Appendix A for the final questionnaire questions and their original constructs. Finally, our preliminary questionnaire included a set of 47 questions. Twenty-seven questions were institutional-related, and 20 questions were social-related.

#### 3.1. Experimental Design and Recruitment

To evaluate our scale, we recruited participants via Amazon Mechanical Turk (AMT). Redmiles et al. found that MTurk responses regarding security and privacy aspects can be generalized to a broader population [53]. Our scale was aimed to assess users' perception of an information system, similar to Suh et al. [60]. To ensure generalization, we tested our scale while referring to several systems, but each participant was exposed to one system only. The systems we chose were Facebook, YouTube, and WhatsApp. Two of the systems have a prominent social aspect, which may raise

privacy concerns (Facebook, WhatsApp), and a third system has a smaller social aspect (YouTube), to cover varying systems.

We screened the participants in several ways. They were required to be 18 years of age or older, and to use the systems frequently (approximately at least once a week). From AMT perspective, the participants were based in the U.S., had an approval rate of 95% or greater, and had at least 100 HITs approved. As we intended to do exploratory factor analysis, we recruited 300 participants. Bryant and Yarnold suggested a minimum ratio of 1:5 of participants per items to conduct EFA [11]. Our questionnaire included 47 items. Thus we assumed we would have at least the desired minimum if recruiting 300 participants. The participants were randomly assigned to one of the systems only. The questions were presented as statements, and the participants were asked about the extent to which they agree with each statement. We used a seven-point Likert scale, where 1 represented low agreement and 7 represented high agreement. The two sub-scales, institutional-related and social-related, were randomly ordered, and the questions within the subscales were randomly ordered as well. We gave participants an "I do not know" option so that we could determine which questions were problematic. The entire study, including all three experiments, was authorized by the institutional ethics review board (IRB) and occurred between May 2018 and February 2019.

Qualified participants followed a link that randomly assigned each participant to one of the three links to the questionnaire, each referring to one of the systems (Facebook, YouTube, WhatsApp). Following previously developed privacy and usability scales [15, 60], we did not include reversely coded statements. In these scales reversely coded questions are rare due to the added complexity they add to the scale. The questionnaire was built using the Qualtrics commercial web survey service. The participants completed an IRB-approved consent form on participation limitations. The mean completion duration was approximately 6.5 minutes, and we paid $0.4 per assignment completion.

Similar to the methods used by Egelman and Peer [17], we took two steps to mitigate social desirability bias on participants' responses, in which some participants may answer questions according to what they believe to be viewed as favorably by others [14]. First, we did not mention "privacy" during recruitment to minimize selection bias. Second, we asked all participants to complete the 10-item Strahan-Gerbasi version of the Marlowe-Crowne Social Desirability Scale [57], which we then correlated with participants' responses to our survey questions. We checked for the existence of straight-lining behavior, in which a participant answers the same answer for all the questions, generally considered to point at superficial thinking [70]. Lastly, following Goodman et al.'s [48] study on AMT, we phrased screening questions to identify participants who would not follow the survey's instructions. If participants failed to answer both questions incorrectly, we excluded their records. After filtering out participants who completed the screening task incorrectly (n = 59) and checking for straight lining behavior (n = 0), we were left with 241 valid responses. See Appendix C for the screening task questions. See Appendix B for the participants' age and gender distribution. The group size of each system was: Facebook: 67, WhatsApp: 78, YouTube: 96.

### 3.2. Results

We performed Pearson correlations between the Strahan-Gerbasi social desirability scale and each question. Except for one question, the observed Pearson's r values corresponded to less than 5% common variance. The remaining question corresponded to less than 10% common variance, the cutoff of which one relationship represents practical importance. Therefore, we chose to treat all the questions as lacking social desirability bias. The result suggests that participants answered truthfully and consistently.

We proceeded to perform Exploratory Factor Analysis (EFA) using Promax rotation to determine which questions should remain in the final questionnaire [58]. We performed four EFAs: one analysis included all the systems and three others for each system separately. We used a loading value of 0.5 as a cut off to include the item within the questionnaire, similar to Egelman and Peer [17]. The number of factors we extracted for each EFA was based on Principal Component Analysis (PCA), using a parallel analysis [2]. For all-systems and Facebook EFAs we extracted four factors, for WhatsApp and YouTube we extracted three and two factors, respectively. First, we removed questions that were below the cutoff value in the EFA that referred to all the systems (7 items). Next, we removed the questions that were below the cutoff in the EFA of each specific system (12 items). We looked for questions that will fit as much as possible to varying types of systems. Therefore, if a question was not good enough for a certain type of system, but was with an appropriate loading value in the other systems, we chose to eliminate it.

Lastly, we re-run the EFA with the remaining 28 questions using the responses of all the systems, ensuring that all the items' loadings are above the cutoff. At this point, we extracted three factors according to the parallel analysis and this analysis resulted also in a 28 items questionnaire. We also checked that none of the final questions was extremely problematic regarding the number of participants choosing "I do not know." Among all the questions (47 items) the highest rate of the N/As responses was 9.5%, and the mean rate was 3.5%. Among the final set of questions, the highest

rate was 6.2% and the mean rate was 3.6% Therefore, we kept all 28 questions.

Finally, our analyses yielded three factors, which we named as institutional, social and risk, partially confirming our hypothesis. Our results showed that users' perceived privacy consist of three distinct aspects, and not only of institutional and social aspects. The questions of the institutional factor are taken from our initial institutional-related section. Respectively, the questions of the social factor are taken from our initial social-related section. However, the questions of the risk factor are mixed of the two original sections, and they are all referring to risk or information sensitivity.

## 4. Finalizing the Scale

We recruited an additional cohort of participants so that we could perform a Confirmatory Factor Analysis (CFA) [58] on the reduced questionnaire. The participants had answered 33 items questionnaire, based on EFA using Varimax rotation. Further Promax rotation eventually reduced the number to of questions to 27.

### 4.1. Method and Demographics

We recruited 300 new participants to respond to the set of the chosen questions. Since at this stage we aimed to have a final scale, we preferred to have more participants per items. Therefore we recruited 300 participants, despite the reduction of the total items number. Following our preliminary results, we removed the Strahan-Gerbasi scale. We kept our screening questions to allow us the removal of suspicious careless responses. We removed the option to answer "I do not know." The course of the experiment was similar to the former experiment. The mean completion duration was approximately 4.23 minutes, and we paid $0.4 per assignment completion. After filtering out participants who completed the screening task incorrectly (n = 82) and checking for straight lining behavior (n = 4), we were left with 214 valid responses. See Appendix B for the participants' age and gender distribution. The group size of each system was: Facebook: 89, WhatsApp: 52, YouTube: 73.

### 4.2. Results

In the following section, we describe several heuristics aimed to explore our scale validity [58]. First analyses are aimed to ensure constructs validity, using PCA, EFA, CFA, as well as convergent and discriminant validity. Next, we performed a reliability analysis, using Cronbach's alpha analysis. The constructs and reliability analyses were conducted based on all the responses (n = 214). Lastly, we analyzed the scale sensitivity, in which we compared the three systems. The sensitivity analysis resulted in changing some of the questions wordings, as we describe in the coming paragraphs.

5.2.1 Constructs Validity. First, we performed a PCA using parallel analysis to extract the number of factors [2]. The scree plot pointed to three factors, as expected. Next, we performed an EFA using Promax rotation and considered an item to be loaded on a factor if its loading exceeded 0.5. The factors and the questions within them were the same as in the preliminary scale. Therefore, all the 27 questions remain within the final scale. The three factors that we extracted predicted 56.1% of the variance. The themes of the factors remained the same: institutional, risk and social. Each of these factors accounted for 25.7%, 16%, and 14.4% of the variance, respectively.

Next, to validate our EFA results, we performed a CFA and examined the model's goodness-of-fit. Multiple popular metrics showed that our data supported the model. Our relative chi-square statistic, $\chi 2/df$, was 2.0. There is no consensus regarding an acceptable cutoff for the ratio, and recommendations range from 5.0 to 2.0. Therefore, our result is acceptable [26]. Our analysis yielded Root Mean Square Error of Approximation (RMSEA) of 0.068 and a Standardized Root Mean Square Residual (SRMR) of 0.065, which is following the recommended maximum cutoff point of 0.08 for both measures [26]; Finally, our Comparative Fit Index (CFI) was 0.92 and Tucker-Lewis Index (TLI) was 0.91, which are above the recommended cutoff of 0.90 [45].

Finally, we conducted convergent and discriminant validity tests. Convergent validity ensures sufficient inter-correlation between each of the construct's variables, while discriminant validity ensures that the constructs are distinct [58]. We found that the average variance extracted (AVE) of each factor is above the acceptable cut-off of 0.5, pointing to convergent validity [24]. As per discriminant validity, we found that the square root of the AVE of each construct was greater than the correlations between the construct and the other constructs in the model [24]. The results are summarized in Appendix E.

*4.2.2 Reliability Analysis*. We examined the scale reliability using Cronbach's alpha. The computed Cronbach's alpha for the full scale was 0.95. Next, all of subscales had excellent internal consistency as well (> 0.9) [22]: institutional: $\alpha$ = 0.95, social: $\alpha$ = 0.9, and risk: $\alpha$ = 0.9. Thus, we concluded that our full scale and the sub-scales each had high reliability.

*4.2.3 Scale Sensitivity Analysis*. Lastly, we compared the systems using the new scale, to have a preliminary notion whether the scale is sensitive enough to detect differences in perceived privacy between systems, similar to the approach taken by Suh et al. [60]. First, we averaged each scale per participant, so each participant now had three scores (institutional, social, risk). Next, we performed Analysis of Variance (ANOVA) per each sub-scale, in which we tested whether there is a significant difference between the sys-

tems. We performed a Tukey post-hoc analysis to find between which systems the difference in the mean score of the scale was significant. The results are summarized in table 1. We can see that for both scales, institutional and risk, there were significant differences between some of the systems. We were surprised by the results, since we would aspect to see a difference in the social aspect primarily, at least between Facebook and YouTube or WhatsApp and YouTube since we chose the systems based on their social aspect.

The ANOVA and the Tukey analyses results brought us to reconsider the statements wordings. The social statements were completely developed and phrased by us, while we considered the previous literature in mind. The results had brought us to notice that we did not include the specified name of the relevant system almost in all social questions, unlike in the other sub-scales questions, which we only slightly modified previously developed questions. Therefore, we added the specified name of the system for those statements as well. To conclude, we see that the survey was sensitive to an extent at this point, detecting some differences between different systems, before finalizing the social statements wordings. See Appendix A for the final suggested scale.

**Table 1. Comparing the systems (Facebook, WhatsApp, and YouTube), exploring in which subscales there are significant differences in the mean score.**

| | ANOVA | Systems compared | Adj. *p*-value (Tukey) |
|---|---|---|---|
| **Institu.** | **$F(2,211) = 5.09, p < 0.01$** | **WA-FB** | **0.007** |
| | | YT-FB | 0.09 |
| | | YT-WA | 0.52 |
| Social | $F(2,211) = 1.69, p = 0.19$ | WA-FB | 0.67 |
| | | YT-FB | 0.492 |
| | | YT-WA | 0.168 |
| **Risk** | **$F(2,211) = 8.63, p < 0.01$** | **WA-FB** | **0.001** |
| | | **YT-FB** | **0.002** |
| | | YT-WA | 0.885 |

## 5. Controlled Experiment and Using the Scale

In the previous sections, we described the development and the steps we took to ensure the internal validation of our scale. In this section, we describe how we used the scale to answer our research questions referring to the applicability of controlled experiment to privacy purposes evaluation, and in which circumstance it can be applied. Unlike as with the previous sections, in which we compared between real systems, and therefore were unable to control for different variables related to privacy design, in this experiment we created scenarios and controlled the desired variables.

### 5.1. Method

To answer our research questions, we designed a between-subject user study, using an online experiment that included a scenario presentation followed by the UPSP scale. We created five scenarios, and per each scenario we created two cases, differing in their privacy design: privacy intrusive design versus privacy respective design. Altogether, we had five background scenarios and ten cases. We recruited 1,026 participants, and they were randomly assigned to one of the ten scenario-case combinations only. We used G*power to estimate the required sample size for T-test analyses and found that the required sample size is 88 participants per group (effect size $d = 0.5$, $\alpha = 0.05$, $1-\beta = 0.95$) [18]. Therefore, we recruited 100 participants per case, and also run several pilots to make sure that the experiments work well, eventually recruited 1,026 participants. Screening parameters for recruiting participants were similar to previous experiments (Sections 3 and 4), except they were not required to be Facebook, WhatsApp or YouTube frequent users. In this experiment, we changed the screening task by shortening the paragraph the participants were required to read (Appendix D).

The background scenarios were developed based on similar principles of previously real privacy case studies that had occurred. For example, one of the scenarios was similar to WhatsApp status update and referred to privacy concerns that were raised as a result of launching the feature [3, 62, 64]. As for the visualization perspective, we designed the general scenarios and the cases based on Ayalon and Toch study [5]. They found that when presenting the privacy characteristics of a system, there is a need to show the human aspect of the problem, rather than presenting it only as a matter of data flow. Qualified participants were first presented with a general explanation, in which the participants were informed that they are about to read a description of a future app and that they are asked to imagine themselves as users in the specific scenario. Next, the participants were presented with the case details, which was consisted of four information sections: 1) *App Presentation* – the app's name followed by a very short description. If required, additional information about the app was provided; 2) *App demonstration* - screenshot, one or more, demonstrating some of the app's interfaces; 3) *Feature presentation* (optional) – in case of a feature within an app, specific information about the feature was provided; 4) *Case description* - description of the specific case and a relevant screenshot, one or more. Lastly, the participants were presented with the

UPSP scale questions. The statements were presented as three sub-scales: institutional, social, and risk. The subscales were ordered accordingly, and the statements within each subscale were randomly ordered.

As we were interested in testing the different privacy aspects of information systems, three background scenarios had a prominent social aspect, and two background scenarios were focused on the institutional aspect. The applications' names that were used as the background scenarios were invented, but we have based the applications' functionalities on existing applications. The three social applications and features used were: 1) iFindRest, which helps the users with finding restaurants based on their location and reserving a table; 2) Message4All app, Tale feature. The app is a messenger app, and the feature enables the users to show content to all the apps' users who have the user's phone number, for a limited time; 3) Message4All app, focusing on groups' details disclosure. Users can view their contacts' shared and non-shared groups. The remaining two institutional applications used were: 4) iFit, a fitness app which helps the users with doing exercises; 5) Message4All app, ads publications, in which ad appears in the chat interface. See Appendix F to view the different scenarios and the two cases per each scenario.

Taking all the participants' responses across the scenarios, the mean completion duration was approximately 6.7 minutes, and we paid an average of $0.63 per assignment completion. After filtering out participants who completed the screening task incorrectly (n = 160) and checking for straight lining behavior (n = 8), we were left with 858 valid responses. See Appendix B for the participants' age and gender distribution. The group size of each scenario-case combination was: iFindRest: intrusive 96, respective 76; Message4All - Tale: intrusive 80, respective 76; Message4All - Groups: intrusive 99, respective 77; iFit: intrusive 86, respective 79; Message4All - Ad: intrusive 87, respective 102.

#### 5.2. Results

We began with re-validating our scale using CFA. Based on the entire sample (n = 858), we examined the model's goodness-of-fit using the same fit statistics as previously and found that our data supported the model: $\chi 2/df = 5.29$, RMSEA = 0.071, SRMR = 0.049, CFI = 0.94, TLI = 0.93. Next, as we assured we could use the scale, we turned to compare between the two cases (intrusive vs. respective) per each scenario. First, we averaged each scale per participant to create three distinct scores (institutional, social, risk). We wanted to compare the two cases per each sub-scale. Therefore, we performed T-tests and used Bonferroni correction for multiple comparisons, in which the *p*-values were multiplied by the number of comparisons.

Our results showed that within different scenarios the differences between the cases were significant and are summarized in Table 2. For all the social background scenarios, we found a significant difference between the cases for at least one of the sub-scales (institutional, social, risk). In the scenario that referred to iFindrest we found that the intrusive design was perceived as riskier compared to the respective design (p = 0.045), and we did not find significant differences in the other categories. In the Message4All app that referred to the Tale feature, we found significant differences between the cases for two of the subscales ($p < 0.05$). The privacy respective design was perceived as respective from the institutional and social aspects. Surprisingly, in the Message4All app that referred to groups information disclosure we found that the respective design was considered as riskier compared to the intrusive design ($p = 0.023$). For the other categories, the difference was insignificant. However, in the institutional background scenarios, iFit and Message4All with the ad presentation, we did not find significant differences between the cases for any of the sub-scales. Figure 1 summarizes the mean sub-scales scores of each scenario, comparing the two cases.

**Table 2. Comparing the cases per each scenario, exploring in which subscales there are significant differences in the mean score.**

| Scenario | Sub-scale | Res. | Int. | Adj. *p* value | Cohen's d |
|---|---|---|---|---|---|
| Social background scenarios | | | | | |
| iFindRest | instit. | 3.88 | 3.73 | 1 | 0.10 |
| | social | 4.11 | 4.33 | 0.301 | 0.26 |
| | **risk** | **4.18** | **4.49** | **0.045** | **0.38** |
| Message4-All, Tale | **instit.** | **4.45** | **3.73** | **0.003** | **0.54** |
| | **social** | **4.72** | **4.40** | **0.018** | **0.45** |
| | risk | 4.77 | 4.6 | 0.32 | 0.26 |
| Message4-All, Group | instit. | 4.04 | 3.76 | 0.689 | 0.18 |
| | social | 4.81 | 4.46 | 0.051 | 0.37 |
| | **risk** | **4.97** | **4.64** | **0.023** | **0.41** |
| Institutional background scenarios | | | | | |
| iFit | instit. | 3.34 | 3.07 | 0.756 | 0.18 |
| | social | 3.84 | 3.84 | 1 | 0.0 |
| | risk | 4.05 | 4.01 | 1 | 0.06 |
| Message4-All, Ad | instit. | 3.67 | 3.33 | 0.523 | 0.20 |
| | social | 4.43 | 4.44 | 1 | 0.01 |
| | risk | 4.68 | 4.77 | 1 | 0.10 |

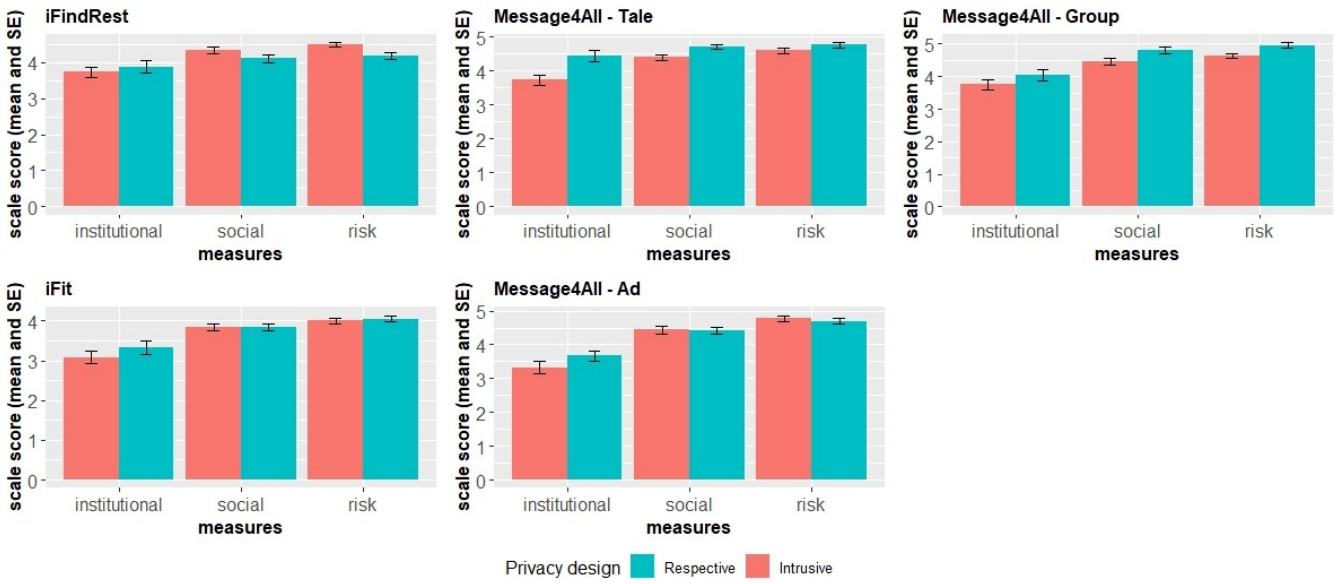



**Figure 1. Per each scenario, we compared the two privacy designs, respective vs. intrusive, per each subscale. For example, in the Message4All - Tale scenario, there were significant differences between the designs for two constructs: institutional and social.**

## 6. Discussion

This study aimed to explore the applicability of controlled experiment methodology to evaluate privacy designs. Towards our exploration, we first attempt to developed a measure to quantify users' privacy perception of a given information system. The scale, in its current form, shows that users perceive information system's privacy via three distinct aspects: institutional, social and risk. This result partially confirms our hypothesis, which referred to institutional and social aspects only. Using our scale, we compared designs which differed in the extent to which they were privacy intrusively designed. Our findings point to a limited ability of controlled experiment methodology to serve as a sensitive way to evaluate privacy design. We saw that the differences between the designs received greater attention when the demonstrated privacy issue had a prominent social aspect, and not, for example, an institutional aspect.

### 6.1. Theoretical Implications

We were motivated by the Privacy by Design (PbD) approach and encouraged by the inclusion of PbD in the European GDPR in 2018. However, PbD can be criticized in a similar way that mainstream system design was criticized by the User Centered Design approach [39]. We argue that ignoring the users and focusing on compliance to regulation will result in systems that are legal but would still make users uncomfortable and go against social norms in particular contexts [46]. Our results point to particular contexts in which system design can be considered as inappropriate. Specifically, our findings suggest that privacy issues with a salient social aspect were highly prone to criticism and observation by the users, compared to the institutional aspect.

Involving the users in the design process may lead to surprising, sometimes even paradoxical, results. In the Message4All scenario, the design that included a message that reminded users that they can disable the disclosure of sensitive information was considered riskier than the alternative design that did not included a message (but in which the sensitive information was collected). Knijnenburg and Kobsa reported on similar results in which messages that were aimed to justify information disclosure decreased the users' trust and satisfaction of the tested system [29]. This result highlights the need to involve the users, showing that the designers, in this case the papers' authors, cannot fully estimate users' perceptions and understandings without asking them directly.

Our findings highlight the promises, and limitations, of our methodology. Controlled experiment methodology is widely used today to provide a fast and affordable evaluation of computing systems. The widespread deployment of this methodology demonstrates that some aspects of user-centered design (UCD) approach are becoming well accepted by today's computing systems' developers, designers and anyone who is part of the decision-making process.

Investigating the applicability of the scale to privacy design evaluation revealed a more complicated picture, in which we saw a significant difference between the cases only in some of the background scenarios. There are several possible ex-

planations for the different results between institutional and social scenarios. One of the explanations can be the difference between the systems' *privacy affordances* [37, 40, 55]. General perceived affordances refer to both the perceived and actual properties of a certain "thing" that define how it can be used [47]. Referring to privacy, previous studies referred to privacy affordances in several contexts. For example, Kou et al. found that Facebook's features as chatrooms and posts' privacy settings affect the users' self-presentation behavior [37]. Liebling and Preibusch suggested to improve gaze tracker by adding privacy affordances to increase the users' privacy [40]. In the current paper we refer to privacy affordances as the ease of the users' ability to understand or foresee the possible consequence of a given privacy issue.

Privacy affordances, as raised in our results, can add another perspective to the privacy paradox debate. The Privacy Paradox is a term usually referring to the gap between people's stated privacy concerns (high) and their actual behavior (disclosing a large amount of information, for example) [33]. Many studies are exploring the paradox, some suggesting possible explanations. One type of explanations refers to the users' constraints of bounded rationality and incomplete information [1], and information asymmetries [12]. These explanations are referring to the users' limited knowledge of the possible consequences of their privacy-related behavior. Our results support these explanations, pointing to different privacy affordances in different types of privacy aspects. For social aspects, privacy affordances are straightforward allowing users to easily imagine possible consequences. As users are actively engaging with social applications, serving as both publishers and audience, users understand what could be the results of posting information to their entire contact list. On the other hand, as with institutional aspects, privacy affordances are much weaker. It is harder to understand the complicated information flows that are behind the way contemporary platforms collect and process their personal information, and which other unknown institutions might access their information and use it as well.

Methodological explanations to the sensitivity of the scale are possible as well. First, the experiment consisted of no more than five scenarios. Possibly, the designs of the institutional scenarios (Message4All – Ad, and iFit) were not sufficiently different surface the problematic privacy issues they ought to represent. Perhaps, if we had used other institutional scenarios we would have received different results. Second, although the scale was validated for its reliability using several acceptable methods, further exploration and improvement is required. Performing EFA had brought us to conclude that there are three distinct constructs. However, it is possible that the difference between the construct "risk" and the two other constructs is not big enough, thus influencing on the ability to differentiate between the privacy designs.

### 6.2. Using the Scale and Design Implications

In this study, we have attempted to develop a scale to measure systems' privacy. Although the scale was designed to capture the users' perceived privacy of a specific system, without limiting the type or the context of the system, the results point to the scale's partial success in fulfilling its intended role. We suggest possible implementations of the scale, however, not without mentioning its limitations to differentiate between all privacy designs. Future implementation of the scale should consider its possible inability to differentiate between privacy designs that are lacking of social aspect.

Following our first suggested explanation, privacy affordances, beyond its theoretical contribution, it also has practical implications. The UPSP scale aims to provide knowledge about the users' perceptions of a system's privacy. Finding significant differences between the designs can point to a good usage of privacy affordances while lacking differences can highlight that the users might not fully understand the possible privacy consequences. Systems' developers should not necessarily give themselves a pat on the back when they do not find a significant difference between the system's privacy designs. They should first look at the score, whether pointing to a high sense of perceived risk, for example. In addition, if in both cases perceived risk is high, but they do not significantly differ, the developers should consider the option the users simply cannot imagine what might be the results of their privacy behavior.

Controlled experiments provide practitioners with new knowledge, for example, which design resulted in a higher conversion rate [32]. Using the UPSP scale provides new knowledge as well. The novelty of our scale is its multi-facets, covering distinct privacy aspects (social, institutional and risk), and its approach, aimed to evaluate systems, rather than individuals' attitudes, as their general privacy concerns. While considering the scale's current uncertain ability to differentiate between privacy designs with a prominent institutional aspect, information system's developers can benefit from using our scale in several ways. First, the scale itself, resulting in three distinct scores per each tested design. A controlled experiment on its own will not provide the required understanding. For example, if the conversion rate was used as a measure, the developers would still lack the knowledge of what was wrong, or right, as perceived by the users. Second, the scale brings the users' perceptions, which might differ from the developers' perception and even from privacy experts. This is similar to other fields as user experience, user interface, usability, and others. Experts are required to set the hypotheses, but the users will eventually determine whether to confirm or reject them. Third, in their study Spiekermann and Cranor suggested guidelines for building privacy-friendly systems, distinguishing between

"privacy-by-policy" versus "privacy-by-architecture." [54] Our study results suggest adding more spheres that should be considered, especially with the rise of social privacy aspect since their study was conducted.

The last implication for design is our structured suggested framework for evaluating users' perceived privacy, as was described in section 5.1. The framework is necessary to demonstrate privacy issues simply and concisely, and yet, understandable by the general population. The process includes five steps: general scenario level: 1) App Presentation; 2) App demonstration; 3) Feature presentation (optional); different versions level: 4) Case description. 5) Lastly, answering the UPSP scale.

### 6.3. Limitations and Future Work

Our study is subject to several limitations that impact its applicability for design and research. First, to evaluate our scale we used five background scenarios. While we have strived to base the scenarios on typical privacy designs, further studies and practical experience are necessary to evaluate it the real world. Second, the participants reflected their opinion about the presented scenario. Their actual behavior in the context of a similar incident might differ, possibly reflecting a weaker difference between the cases. Third, as norms around privacy evolve these days quickly, the scale should be continuously evaluated to see that it reflect contemporary notions. Lastly, as we have suggested a method to evaluate privacy designs, the study population should be sampled and adjusted to particular systems and scenarios. As with many privacy studies, the use of Mechanical Turk as the study's population may not reflect the actual demographics of the intended system.

Based on the study results we are developing *A/P(rivacy) Testing*[1], a platform that will enable other researchers and developers to use our scale and to compare privacy designs easily. Future studies can explore real systems or focus on specific challenges, for example, exploring different ways to visualize consent form and the visualization's effect on users' perceived privacy.

## 7. Acknowledgment

This work was supported by the Shulamit Aloni Scholarship by the Israeli Ministry of Science and Technology, grant number 314575, and by the ICRC – Blavatnik Interdisciplinary Cyber Research Center, grant number 590713. We would also like to thank Luiza Jarovsky for helping us with finalizing the scale and Shany Peter for developing the A/P Testing tool.

## References

[1] Acquisti, A. and Grossklags, J. 2005. Privacy and rationality in individual decision making. *IEEE Security and Privacy*. 3, 1 (2005), 26–33.

[2] Ahmad, sarah sabir 1999. Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*. 4, 3 (1999), 272.

[3] As Facebook Raised a Privacy Wall, It Carved an Opening for Tech Giants: 2018. *https://www.nytimes.com/2018/12/18/technology/facebook-privacy.html?module=inline*.

[4] Awad, N. and Krishnan, M. 2006. The personalization privacy paradox: an empirical evaluation of information transparency and the willingness to be profiled online for personalization. *Mis Quarterly*. 30, 1 (2006), 13–28.

[5] Ayalon, O. and Toch, E. 2018. Crowdsourcing Privacy Design Critique : An Empirical Evaluation of Framing Effects. *Submitted*. (2018).

[6] Ayalon, O., Toch, E., Hadar, I. and Birnhack, M. 2017. How Developers Make Design Decisions about Users' Privacy: The Place of Professional Communities and Organizational Climate. *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW '17 Companion*. (2017), 135–138.

[7] Balebako, R., Marsh, A., Lin, J., Hong, J.I., Cranor, L.F. and Faith Cranor, L. 2014. The Privacy and Security Behaviors of Smartphone App Developers. *Workshop on Usable Security (USEC)*. (2014).

[8] Bamberger, K.A. and Mulligan, D.K. 2011. *Privacy on the Books and on the Ground.* MIT Press.

[9] Bechmann, A. 2014. Non-Informed Consent Cultures: Privacy Policies and App Contracts on Facebook. *Journal of Media Business Studies*. 11, 1 (2014), 21–38.

[10] Boyd, D. 2010. Making Sense of Privacy and Publicity. *South by Southwest (SXSW 2010)– transcription of the talk*.

[11] Bryant, F.. B. and Yarnold, P.. R. 1995. Principal-components analysis and exploratory and confirmatory factor analysis. *Reading and understanding multivariate statistics*. L.G. Grimm and P.. R. Yarnold, eds. American Psychological Association. 99–136.

[12] Buck, C., Horbel, C., Germelmann, C.C. and Eymann, T. 2014. The unconscious app consumer: Discovering and comparing the information-seeking patterns among mobile application consumers. *European Conference on Information Systems (ECIS)* (2014).

[13] Cavoukian, A. 2009. Privacy by design: The 7

[1] www.aprivacytesting.com

foundational principles. Information and Privacy Commissioner of Ontario, Canada.

[14] Crowne, D.P. and Marlowe, D. 1960. A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*. 24, 4 (1960), 349–354.

[15] Dinev, T., Xu, H., Smith, J.H. and Hart, P. 2013. Information privacy and correlates: An empirical attempt to bridge and distinguish privacyrelated concepts. *European Journal of Information Systems*. 22, 3 (2013), 295–316.

[16] Egelman, S., Felt, A.P. and Wagner, D. 2013. Choice architecture and smartphone privacy: There's a price for that. *The Economics of Information Security and Privacy*. (2013), 211–236.

[17] Egelman, S. and Peer, E. 2015. Scaling the Security Wall: Developing a Security Behavior Intentions Scale (SeBIS). *Proceedings of the ACM CHI'15 Conference on Human Factors in Computing Systems*. 1, (2015), 2873–2882.

[18] Faul, F., Erdfelder, E., Lang, A.G. and Buchner, A. 2007. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*. 39, 2 (2007), 175–191.

[19] Featherman, M.S. and Pavlou, P.A. 2003. Predicting e-services adoption: A perceived risk facets perspective. *International Journal of Human Computer Studies*. 59, 4 (2003), 451–474.

[20] Felt, A.P., Egelman, S. and Wagner, D. 2012. I've got 99 problems, but vibration ain't one. *Proceedings of the second ACM workshop on Security and privacy in smartphones and mobile devices - SPSM '12*. (2012), 33.

[21] GDPR: *https://gdpr-info.eu/art-25-gdpr/*. Accessed: 2018-01-16.

[22] George, D. and Mallery, P. 1999. *SPSS for Windows Step by Step: A simple guide and reference*. Needham Heights, MA: Allyn & Bacon.

[23] Hadar, I., Hasson, T., Ayalon, O., Toch, E., Birnhack, M., Sherman, S. and Balissa, A. 2017. Privacy by designers: software developers' privacy mindset. *Empirical Software Engineering*. (2017), 1–31.

[24] Hair, J.F., Black, W.C., Babin, B.J. and Anderson, R.E. 2014. *Pearson New International Edition: Multivariat Data Analysis*.

[25] Hong, W. and Thong, J.Y.L. 2013. Internet privacy concerns: an integrated conceptualization and four empirical studies. *Mis Quarterly*. 37, 1 (2013), 1–3.

[26] Hooper, D., Coughlan, J., Mullen, M.R., Mullen, J., Hooper, D., Coughlan, J. and Mullen, M.R. 2008. " Structural Equation Modelling: Guidelines for Determining Model Fit Structural Equation Modelling: Guidelines for Determining Model Fit. *The Electronic Journal of Business Research Methods*. 6, 1 (2008), 53–60.

[27] ICO (Information Commissioner's Office) 2014. Conducting privacy impact assessments code of practice. *Ico.Org.Uk*. (2014), 1–55.

[28] Jarvenpaa, S.L., Tractinsky, N. and Saarinen, L. 1999. Consumer Trust in an Internet Store: a Cross-Cultural Validation. *Journal of Computer-Mediated Communication*. 5, 2 (1999).

[29] Knijnenburg, B.P.B. and Kobsa, A. 2013. Making decisions about privacy: Information disclosure in context-aware recommender systems. *ACM Transactions on Interactive Intelligent Systems*. 3, 3 (2013), 20:1–20:23.

[30] Kobsa, A., Hichang, C. and Knijnenburg, B.P. 2016. The Effect of Personalization Provider Characteristics on Privacy Attitudes and Behaviors: An Elaboration Likelihood Model Approach Alfred. *Journal of the Association for Information Science and Technology*. 67, 11 (2016), 2587–2606.

[31] Kohavi, R., Deng, A., Frasca, B., Walker, T., Xu, Y. and Pohlmann, N. 2013. Online controlled experiments at large scale. *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '13*. (2013), 1168.

[32] Kohavi, R., Longbotham, R., Sommerfield, D. and Henne, R.M. 2009. Controlled experiments on the web: Survey and practical guide. *Data Mining and Knowledge Discovery*. 18, 1 (2009), 140–181.

[33] Kokolakis, S. 2017. Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers and Security*. 64, (2017), 122–134.

[34] Koops, B.J. and Leenes, R. 2014. Privacy regulation cannot be hardcoded. A critical comment on the "privacy by design" provision in data-protection law. *International Review of Law, Computers & Technology*. 28, 2 (2014), 159–171.

[35] Krasnova, H., Günther, O., Spiekermann, S. and Koroleva, K. 2009. Privacy concerns and identity in online social networks. *Identity in the Information Society*. 2, 1 (2009), 39–63.

[36] Krasnova, H., Spiekermann, S., Koroleva, K. and Hildebrand, T. 2010. Online social networks: Why we disclose. *Journal of Information Technology*. 25, 2 (2010), 109–125.

[37] Kuo, F.-Y., Tseng, C.-Y., Tseng, F.-C. and Lin, C.S.

2013. A Study of Social Information Control Affordances and Gender Difference in Facebook Self-Presentation. *Cyberpsychology, Behavior, and Social Networking*. 16, 9 (2013), 635–644.

[38] Langheinrich, M. 2001. Privacy by Design - Principles of Privacy-Aware Ubiquitous Systems. *3rd international conference on Ubiquitous Computing*. (2001), 273–291.

[39] Law, E.L.-C., Roto, V., Hassenzahl, M., Vermeeren, A.P.O.S. and Kort, J. 2009. Understanding, scoping and defining user experience. *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*. June 2014 (2009), 719.

[40] Liebling, D.J. 2014. Privacy Considerations for a Pervasive Eye Tracking World. (2014), 1169–1177.

[41] Lin, J., Sadeh, N., Amini, S., Lindqvist, J., Hong, J.I. and Zhang, J. 2012. Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing. *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12*. (2012), 501.

[42] Malhotra, N.K., Kim, S.S. and Agarwal, J. 2004. Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information Systems Research*. 15, 4 (2004), 336–355.

[43] Metzger, M.J. 2004. Privacy, Trust, and Disclosure: Exploring Barriers to Electronic Commerce. *Journal of Computer-Mediated Communication*. 9, 4 (2004).

[44] Mohamed, N. and Ahmad, I.H. 2012. Information privacy concerns, antecedents and privacy measure use in social networking sites: Evidence from Malaysia. *Computers in Human Behavior*. 28, 6 (2012), 2366–2375.

[45] Netemeyer, R.G., Bearden, W.O. and Subhash, S. 2003. *Scaling procedures: Issues and applications*. Sage Publications.

[46] Nissenbaum, H. 2010. *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press.

[47] Norman, D. 2013. *The design of everyday things: Revised and expanded edition*. Basic books.

[48] Oppenheimer, D.M., Meyvis, T. and Davidenko, N. 2009. Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*. 45, 4 (2009), 867–872.

[49] Page, X., Tang, K., Stutzman, F. and Lampinen, A. 2013. Measuring networked social privacy. *Proceedings of the 2013 conference on Computer supported cooperative work companion - CSCW '13*. (2013), 315–320.

[50] Poikela, M. and Toch, E. 2017. Understanding the Valuation of Location Privacy: a Crowdsourcing-Based Approach. *Proceedings of the 50th Annual Hawaii International Conference on System Sciences*. (2017), 1985–1994.

[51] Quinn, K. and Epstein, D. 2018. #MyPrivacy: How Users Think About Social Media Privacy. *Proceedings of the 9th International Conference on Social Media and Society - SMSociety '18*. (2018), 360–364.

[52] Raynes-Goldie, K. 2010. Aliases, creeping, and wall cleaning: Understanding privacy in the age of Facebook. *First Monday*. 15, 1 (2010).

[53] Redmiles, E.M., Kross, S., Pradhan, A. and Mazurek, M.L. 2017. How Well Do My Results Generalize? Comparing Security and Privacy Survey Results from MTurk and Web Panels to the U.S. *University of Maryland Technical Reports of the Computer Science Department*. (2017).

[54] Spiekermann, S. and Cranor, L.F. 2009. Engineering privacy. *IEEE Transactions on Software Engineering*. 35, 1 (2009), 67–82.

[55] Stark, L. and Tierney, M. 2014. Lockbox : mobility , privacy and values in cloud storage. (2014), 1–13.

[56] Steinbart, P., Keith, M.J. and Babb, J.S. 2017. Measuring Privacy Concerns and the Right to Be Forgotten. (2017), 4967–4976.

[57] Strahan, R. and Gerbasi, K.C. 1972. Short, homogeneous versions of the Marlow-Crowne Social Desirability Scale. *Journal of Clinical Psychology*. 28, 2 (1972), 191–193.

[58] Straub, D., Boudreau, M.-C. and Gefen, D. 2004. Validation Guidelines for Is Positivist. *Communications of the Association for Information Systems*. 13, 24 (2004), 380–427.

[59] Stutzman, F. 2006. An evaluation of identity-sharing behavior in social network communities. *International Digital and Media Arts Journal*. 3, 1 (2006), 10–18.

[60] Suh, H., Shahriaree, N., Hekler, E.B. and Kientz, J.A. 2016. Developing and Validating the User Burden Scale. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*. (2016), 3988–3999.

[61] Vitak, J. 2012. The Impact of Context Collapse and Privacy on Social Network Site Disclosures. *Journal of Broadcasting and Electronic Media*. 56, 4 (2012), 451–470.

[62] WARNING: Google Buzz Has A Huge Privacy

Flaw: 2010. *https://www.businessinsider.com/warning-google-buzz-has-a-huge-privacy-flaw-2010-2*.

[63] Westin, A.F. 1967. *Privacy and Freedom*. New York: Atheneum.

[64] WhatsApp Status Update: Here's Why People Are Scared Of This Feature: 2017. *https://www.ndtv.com/offbeat/whatsapp-status-heres-why-people-are-scared-of-this-feature-1663139*.

[65] Wong, R.Y. and Mulligan, D.K. 2019. Bringing Design to the Privacy Table: Broadening "Design" in "Privacy by Design" Through the Lens of HCI. (2019).

[66] Wong, R.Y. and Mulligan, D.K. Bringing Design to the Privacy Table Broadening " Design " in " Privacy by Design " Through the Lens of HCI.

[67] Xu, H. 2007. The Effects of Self-Construal and Perceived Control on Privacy Concerns. *Twenty Eighth International Conference on Information Systems*. 6, 1 (2007), 1–14.

[68] Xu, H., Teo, H.-H., Tan, B.C.Y. and Agarwal, R. 2010. The Role of Push-Pull Technology in Privacy Calculus: The Case of Location-Based Services. *Journal of Management Information Systems*. 26, 3 (2010), 135–174.

[69] Young, A.L. and Quan-Haase, A. 2013. PRIVACY PROTECTION STRATEGIES ON FACEBOOK: The Internet privacy paradox revisited. *Information Communication and Society*. 16, 4 (2013), 479–500.

[70] Zhang, C. and Conrad, F.G. 2014. Speeding in Web Surveys : The tendency to answer very fast and its association with straightlining. 8, 2 (2014), 127–135.

## 10. Appendix

### A Final Scale

| | Ref. | In. | Institu. | Risk | Social |
|---|---|---|---|---|---|
| I think I have control over what personal information is shared by [X] with other companies. | [67] | Ct | 0.97 | 0.07 | -0.14 |
| I believe I have control over how my personal information is used by [X]. | | | 0.79 | 0.01 | 0.03 |
| I believe I have control over what personal information is collected by [X]. | | | 0.76 | 0.07 | 0.06 |
| It is clear whether my personal information is shared with other companies. | [25] | Su | 0.78 | 0.04 | -0.06 |
| I believe that [X] will prevent unauthorized people from accessing my personal information in their databases. | | | 0.54 | -0.15 | 0.21 |
| I believe my personal information is accessible only to those authorized to have access. | | | 0.71 | -0.14 | 0.05 |
| It is clear what information about me [X] keeps in their databases. | [4] | Tr | 0.74 | 0.00 | 0.08 |
| It is clear how long [X] retains my information. | | | 0.77 | 0.16 | -0.02 |
| The purposes for which [X] asks for my information are clear. | | | 0.77 | -0.01 | 0.03 |
| It is clear how [X] uses my personal information. | [25] | | 0.86 | 0.00 | -0.02 |
| I believe that if I would I ask, [X] will allow me to delete my personal information. | [56] | D | 0.60 | 0.04 | 0.14 |
| I think that it will be easy to delete my information from [X]. | | | 0.61 | -0.03 | 0.18 |
| I think it would be risky to give my personal information to [X]. | [19] | R | -0.12 | 0.71 | 0.05 |
| I think that there would be a high potential for privacy loss associated with giving my personal information to [X]. | | | -0.04 | 0.67 | 0.03 |
| My Personal information could be inappropriately used by [X]. | | | -0.26 | 0.57 | 0.08 |
| I think that providing [X] with my personal information would involve many unexpected problems. | | | 0.08 | 0.82 | 0.00 |
| I do not feel comfortable with the type of information I share using [X]. | | | 0.12 | 0.70 | -0.13 |
| Considering the information I provide to [X], and the people who might see it, I think it would be risky to give my personal information to [X]. | | | 0.12 | 0.80 | -0.09 |
| Considering the information I provide to [X], and the people who might see it, I think that there would be a high potential for privacy loss associated with giving my personal information to [X]. | | | 0.00 | 0.70 | 0.01 |
| Considering the information I provide to [X], and the people who might see it, I think that providing [X] with my personal information would involve many unexpected problems. | | | 0.03 | 0.79 | 0.09 |
| I can understand whether people who I may know (friends, family, classmates, colleagues, acquaintances, etc.) have access to my personal information on [X]. | [59] | Id | -0.12 | 0.10 | 0.72 |
| It is clear who is the audience of my shared information on [X]. | | | 0.13 | 0.05 | 0.70 |
| It looks easy to restrict un-intended people from viewing my personal information on [X]. | [69] | Ps | 0.09 | -0.06 | 0.72 |
| It looks easy to manage who can view my personal information on [X]. | | | 0.01 | -0.07 | 0.73 |
| I think [X] allows me to restrict the access to some of my personal information to some people. | | | -0.11 | -0.06 | 0.75 |
| I think I have control over what personal information is shared by [X] with other people. | [67] | Ct | 0.22 | 0.06 | 0.63 |
| It is clear what information about me others can see on [X]. | [25] | Tr | 0.13 | 0.05 | 0.70 |

**Ct**: Perceived information control, **Cf**: confidentiality, **Tr**: Importance of information transparency, **Su**: Secondary usage, **D**: Data deletion, **R**: Perceived privacy risk, **Is**: Information sensitivity, **Ps**: Protection strategies, **Id**: Identity sharing

### B Gender and Age Distribution

| Experiment | N | Gender distribution (%) | | | Age distribution (%) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Female | Male | Did not reveal | 18-24 | 25-34 | 35-44 | 45-54 | 55-64 | 65+ |
| Preliminary scale | 241 | 80 | 158 | 3 | 34 | 138 | 43 | 16 | 9 | 1 |
| Finalizing the scale | 214 | 75 | 139 | | 25 | 101 | 39 | 31 | 10 | 8 |
| Using the scale | 858 | 380 | 471 | 7 | 85 | 366 | 190 | 113 | 71 | 33 |

### C Screening Task – First Two Experiments

Former studies in the field of decision making show that people, when making decisions and answering questions, are not always paying attention and are minimizing their effort as much as possible. A few studies show that over 50% of people don't carefully read questions. If you are reading this paragraph, in the first question please select the box marked 'other' and type 'evaluating information systems is fun' in the box below. Do not select anything else. In the second question, please select 'four'. Thank you for participating and taking the time to read through the questions carefully!

*What was this study about?* [Information systems evaluation, Making decisions about information systems, Evaluating information systems, Other]

*It is common to evaluate information systems*. [Strongly disagree (1), (2), (3), (4), Strongly agree (5)]

### D Screening Task – Third Experiment

A few studies show that over 50% of people don't carefully read questions. If you are reading this paragraph, in the first question please select 'two'. In the second question, please select 'four'. Thank you for participating and taking the time to read through the questions carefully!

*I usually take the time to evaluate information systems.* [Strongly disagree (1), (2), (3), (4), Strongly agree (5)]

*I think that evaluating information systems is important. [Strongly disagree (1), (2), (3), (4), Strongly agree (5)]*

### E Internal consistency and discriminant validity of constructs

| | | | | Factors correlations | | |
|---|---|---|---|---|---|---|
| | Cronbach's α | AVE | SQRT(AVE) | Institutional | Social | Risk |
| Institutional | 0.95 | 0.56 | 0.75 | | 0.63 | -0.23 |
| Social | 0.9 | 0.53 | 0.72 | | | -0.29 |
| Risk | 0.9 | 0.53 | 0.73 | | | |

### F Controlled Experiment: General Scenario Followed by One of the Two Cases

**General scenario presentation: iFindRest**

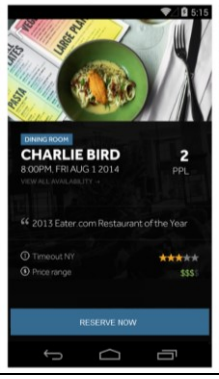
You are presented with a description of a future app, and we ask that you imagine yourself as a user in the specific scenario. Please read the description carefully and answer the following questions.

**iFindRest**

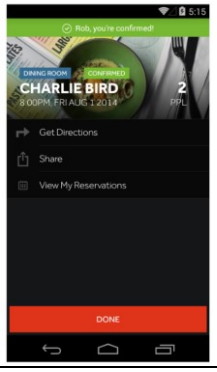
iFindRest is an app that helps with finding restaurants based on location and reserving a table in the desired restaurant.

The following screenshots demonstrate the app's user interface:

Reserving a table at the desired restaurant, at the desired time




Confirmation screen




**Case 1: Privacy protective design**

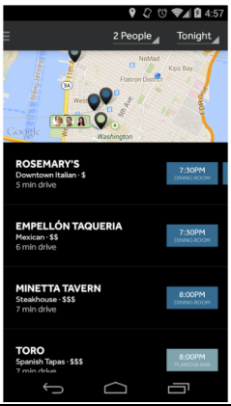
**The scenario**

Imagine that it is around 7:00 PM and you and your friend would like to go for a dinner at a nearby restaurant. You are using iFindRest to look for restaurants in your area, based on your current location. On the screen, you can see relevant restaurants. The restaurant that is marked in green indicates that other users, who are in your contact list, had made reservations to this restaurant at similar hours to yours. You cannot see who these users are since the default choice is not to share their identity publicly with their contact list, and they probably kept it as is.


**Case 2: Privacy intrusive design**

**The scenario**

Imagine that it is around 7:00 PM and you and your friend would like to go for a dinner at a nearby restaurant. You are using iFindRest to look for restaurants in your area, based on your current location. On the screen, you can see relevant restaurants. The restaurant that is marked in green indicates that other users, who are in your contact list, had made reservations to this restaurant at similar hours to yours. You can also see who these users are since the default choice is to share their identity publicly with their contact list, and they probably kept it as is.



## General scenario presentation: iFit

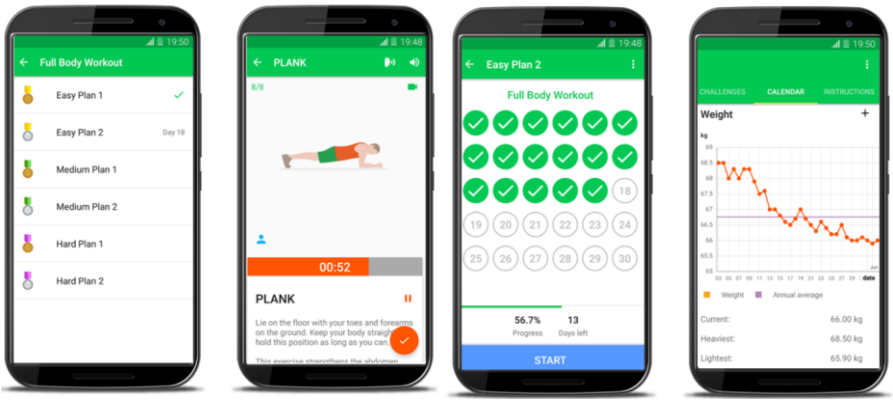
You are presented with a description of a future app, and we ask that you imagine yourself as a user in the specific scenario. Please read the description carefully and answer the following questions.

**iFit**

iFit is a fitness app which helps the users with doing exercises. The app provides a 30 days training programs for different parts of the body, at different difficulty levels.

The following screenshots demonstrate the app's user interface:

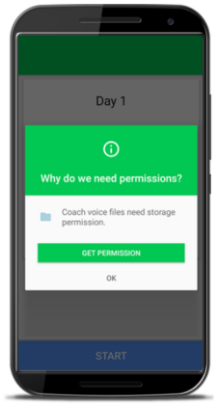
Choosing the desired difficulty level | Doing a specific exercise | Process recording | Weight recording



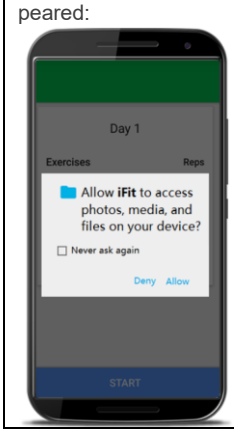

### Case 1: Privacy protective design

**The scenario**

Imagine that this is the first time that you are using iFit. You chose "Easy Plan 1" which focuses on the abdominal muscles. You pressed "start" and the following message appeared:



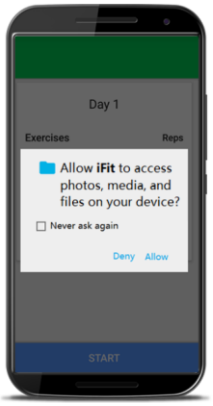

You clicked "GET PERMISSION" and the following message appeared:




### Case 2: Privacy intrusive design

**The scenario**

Imagine that this is the first time that you are using iFit. You chose "Easy Plan 1" which focuses on the abdominal muscles. You pressed "start" and the following message appeared:



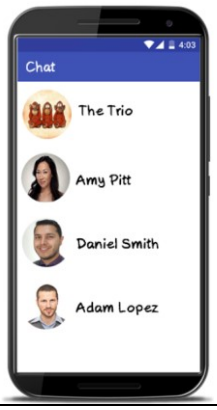
## General scenario presentation: Message4All -Tale

You are presented with a description of a future app, and we ask that you imagine yourself as a user in the specific scenario. Please read the description carefully and answer the following questions.

**Message4All**
Message4All is a messenger app, similar to apps like WhatsApp, Snapchat, etc.

Users can chat with every contact on their phone. However, they are usually using the app for chatting with people with whom they have a close relationship, such as family, friends, and colleagues,
by sending text messages, photos, etc. The app is used for both one-on-one and groups chat conversations.

The following screenshot demonstrate the app's user interface:




## Case 1: Privacy protective design

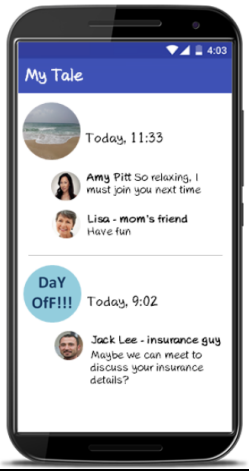
**Tale** is a feature in Message4All that allows the users to show content which can be seen by anyone who has the user's phone number and has Message4All installed. The content will be available for 24 hours only and will be automatically dismissed afterward.

**The scenario**
Imagine that you decided to try the Tale feature. You took a day off and were about to share a video showing the beach you went to. After pressing the "Share" button, the following message appeared on the screen:


## Case 2: Privacy intrusive design

**Tale** is a feature in Message4All that allows the users to show content which can be seen by anyone by default, which has the user's phone number and has Message4All installed. The content will be available for 24 hours only and will be automatically dismissed afterward.

**The scenario**
Imagine that you decided to try the Tale feature. You took a day off and shared two tales. The first one was a text tale and the second contained a video of the beach you went to. During the day, few people commented on your tales, with some of them you rarely speak. You can see your tales and their comments as demonstrated in the following screenshot:



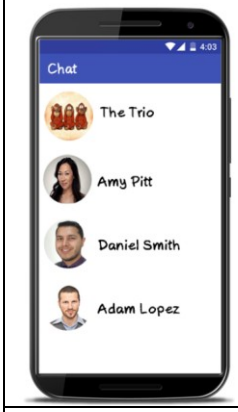
## General scenario presentation: Message4All - Groups

You are presented with a description of a future app, and we ask that you imagine yourself as a user in the specific scenario. Please read the description carefully and answer the following questions.

**Message4All**
Message4All is a messenger app, similar to apps like WhatsApp, Snapchat, etc.

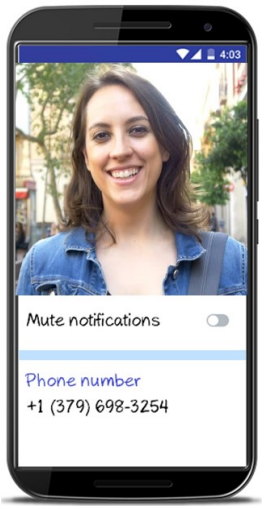
Users can chat with every contact on their phone. However, they are usually using the app for chatting with people with whom they have a close relationship, such as family, friends, and colleagues, by sending text messages, photos, etc. The app is used for both one-on-one and groups chat conversations.

The following screenshot demonstrate the app's user interface:



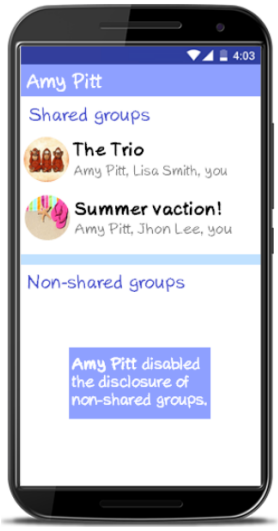

**Contact person details**
Within Message4All contact list, a user can get further information about specific contact person and set settings. For example, the user can review the groups that the contact person is part of, both groups that they have in common and also those they do not share.

The following screenshot demonstrates the app's contact person interface:



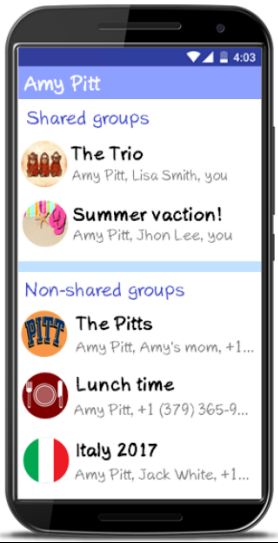

## Case 1: Privacy protective design

**The scenario**
Imagine that you have a friend named Amy Pitt and you often chat with her using Message4All. You wanted to look for a group that you remembered that you are both members of. Therefore, you looked at her details on the app. Here is a screenshot that provides information about Amy's groups.




## Case 2: Privacy intrusive design

**The scenario**
Imagine that you have a friend named Amy Pitt and you often chat with her using Message4All. You wanted to look for a group that you remembered that you are both members of. Therefore, you looked at her details on the app. Here is a screenshot that provides information about Amy's groups.



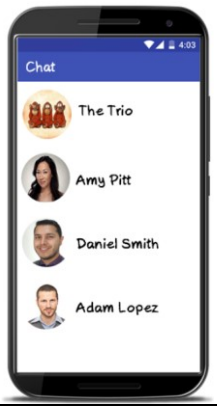
## General scenario presentation: Message4All - Ad

You are presented with a description of a future app, and we ask that you imagine yourself as a user in the specific scenario. Please read the description carefully and answer the following questions.

**Message4All**
Message4All is a messenger app, similar to apps like WhatsApp, Snapchat, etc.

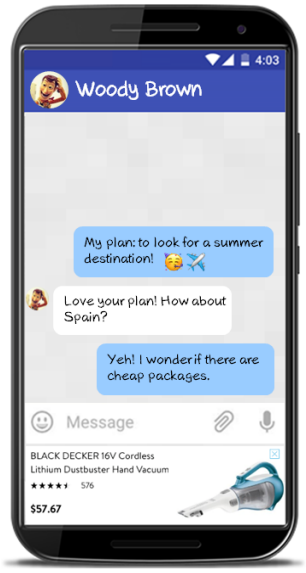
Users can chat with every contact on their phone. However, they are usually using the app for chatting with people with whom they have a close relationship, such as family, friends, and colleagues, by sending text messages, photos, etc. The app is used for both one-on-one and groups chat conversations.

The following screenshot demonstrate the app's user interface:




## Case 1: Privacy protective design

**The scenario**
Imagine that you are using Message4All to chat with your friend Woody. Here is the screenshot of your chat:




## Case 2: Privacy intrusive design

**The scenario**
Imagine that you are using Message4All to chat with your friend Woody. Here is the screenshot of your chat:

# A Self-Report Measure of End-User Security Attitudes (SA-6)

Cori Faklaris, Laura Dabbish and Jason I. Hong
*Human-Computer Interaction Institute, School of Computer Science, Carnegie Mellon University*
*Pittsburgh, PA, USA*
*{cfaklari, dabbish, jasonh}@cs.cmu.edu*

## Abstract

We present SA-6, a six-item scale for assessing people's security attitudes that we developed by following standardized processes for scale development. We identify six scale items based on theoretical and empirical research with sufficient response variance, reliability, and validity in a combined sample ($N$ = 478) from Amazon Mechanical Turk and a university-based study pool. We validate the resulting measure with a U.S. Census-tailored Qualtrics panel ($N$ = 209). SA-6 significantly associates with self-report measures of behavior intention and recent secure behaviors. Our work contributes a lightweight method for (1) quantifying and comparing people's attitudes toward using recommended security tools and practices, and (2) improving predictive modeling of who will adopt security behaviors.

## 1. Introduction

The human in the loop is often the weakest link in any security system [17,78]. Understanding people's attitudes toward security technology is key to designing systems that are both usable and tough to breach. For this reason, a fair amount of research in usable security and privacy employs in-depth interviews and observation with small samples to understand people's attitudes toward a security practice or technology, e.g. [12,20,29,36,73]. However, we need to seek ways to operationalize such concepts in efforts to better understand the phenomenon and its relation with causes and outcomes in a more robust way e.g., experiments and longitudinal surveys. It is not always feasible or appropriate to utilize a qualitative approach. It is time-consuming to identify and label the concepts underlying people's open-ended responses, and such custom analyses are prone to error. We need a quantitative measure in order to systematically assess and compare users' security attitudes.

The current state of the art for measuring users' thinking about security practices is the Security Behavior Intentions Scale [32,33]. SeBIS' 16 items are grounded in security expert recommendations for user behavior in four areas: device securement, updates, password management, and proactive awareness. While SeBIS can tell us the degree to which a user intends to comply with these expert recommendations, it cannot tell us people's attitudes about security behaviors.

A measure of security attitudes supports research on differences in security-related intentions and behaviors. Attitudes represent people's evaluation of objects, groups, events, that is, how they orient to the world around them [4]. An extensive body of research in psychology examines attitudes, their antecedents and consequences, and their relationship to intentions and behavior [4,6,18,49]. In fields as disparate as organizational psychology [57] and environmental sustainability e.g. [9,43], researchers measure attitudes to understand behavior and general tendencies. In security, such a measure would be useful to understand what leads to different security attitudes, and the effect of these attitudes on intentions and behavior.

For this purpose, we introduce a 6-item self-report measure of security attitudes: SA-6. Our measure is based on user-centered empirical and theoretical studies of awareness, motivation to use and knowledge of expert-recommended security tools and practices (*security sensitivity*) [20–24]. Using principles of psychological scale development [28,39,46,53], we generate 48 candidate items that on their face corresponded to prior work on security attitudes and that pilot testers found to be unambiguous and easily answered. Through iterative rounds of analysis, we narrow to six items that demonstrated desired response variance, factor loadings, reliability, and validity using data from Amazon Mechanical Turk and a university-based study pool (combined $N$ = 478) and from a U.S. Census-tailored panel ($N$ = 209).

We find SA-6 to be significantly associated with self-report measures of behavior intention. Using linear regression, we found SA-6 explained 28% of the variance in SeBIS ($p$<.01). This result is consistent with longstanding psychological evidence of the relationship between attitudes and behavior intention [5,7,8,37,69,75]. Our data shows SA-6 also relates with measures of subjective norms, chiefly privacy, and perceived behavior control, such as impulsivity, self-efficacy and internet know-how. Our data also shows SA-6 differs as expected by personal experiences and hearing/seeing reports of security breaches, and by age, gender, education and income level. These results, predicted by the Theory of Reasoned Action Model [37], demonstrate the convergent and discriminant validity of this scale [28,39,46,53].



*USENIX Symposium on Usable Privacy and Security (SOUPS) 2019*. August 11 -- 13, 2019, Santa Clara, CA, USA.

We also find SA-6 significantly associated with self-reported recent security behavior. Here we go beyond previous work on security behavior intentions, connecting attitudes and intentions to people's recalled security actions in the past week. We find that SA-6 and SeBIS associate with recent security actions and support for SeBIS as a partial mediator of SA-6's influence on reported recent security behavior. Our results suggest that SA-6 may help model who is likely to act on security recommendations and who will benefit most from security awareness training or tutorials.

This paper makes the following contributions:

- The introduction of a six-item validated self-report measure of security attitudes (SA-6) for use by researchers and practitioners to systematically assess and compare user attitudes toward security techniques;
- An analysis of the relationship between security attitudes, security intentions, and recent security actions;
- A discussion of how and why to use SA-6 for measuring security attitudes to explain and predict user adoption of recommended security behaviors.

## 2. Related Work

Most human behavior is goal-directed [8]. But for most computer users, staying secure and avoiding relevant threats is a secondary goal at best. The need to understand how to nudge adoption of secure behaviors *in spite of this* underpins much prior work integrating psychology with cybersecurity.

To develop our scale, we identified a concept in the cybersecurity literature that corresponds to the psychological conception of attitude. We then identified concepts that could be expected to relate to and vary with this attitude factor according to theoretical models of how accept and adopt expert-recommended secure tools and practices.

### 2.1. Attitudes

*Attitudes* represent people's evaluation of objects, groups, events, that is how they orient to the world around them [4]. Eagly and Chaiken [30] define an attitude as "a psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor." An extensive body of research in psychology examines attitudes, their antecedents and consequences and their relationship to intentions and behavior [4,6]. In fields as disparate as organizational psychology [57] to environmental sustainability e.g. [9,43], researchers measure attitudes to understand behavior.

To develop our measure of security attitudes, we examined the cybersecurity literature for work that documented user attitudes about expert-recommended tools and practices. The focal concept we identified is *security sensitivity.*

### 2.2. Security Sensitivity

In the field of usable security, end-user *security sensitivity* is defined by Das as "the awareness of, motivation to use, and knowledge of how to use security tools" and practices [20]. Das and collaborators based this construct on empirical findings in interview studies that many people believe themselves in no danger of falling victim to a security breach and are unaware of the existence of tools to protect them against those threats; also, they perceive the inconvenience and cost to their time and attention of using these tools and practices as outweighing the harm of experiencing a security breach; and, they think these measures are too difficult to use or lack the knowledge to use them effectively [20–23].

Das summarized the concept as a series of six questions, which focus in parallel on tools and threats [20]. Restated, these six sub-dimensions are: *awareness of the existence of security threats*; *awareness of the existence of security measures* (tools, behaviors and strategies) that can be used to counteract threats; *motivation to counteract security threats*; *motivation to use security measures* to counteract threats; *knowledge of the relevance of security threats*; and *knowledge of how to use security measures* to counteract relevant threats. This builds in turn on theoretical and empirical work from Davis and others [25,26] on *user perceptions of usefulness and ease of use,* from Egelman et al. [31]'s adaptation of the *Communication-Human Information Processing* model to end-user security, and from Rogers' *Diffusion of Innovations* model [61] of how messages spread in a social network about a new idea.

We used the literature from Das et al. on security sensitivity as a main source of items to test for inclusion in SA-6.

### 2.3. User Acceptance Theories and Models

Davis et al.'s *Technology Acceptance Model* [25,26] (TAM) was among the first to integrate users' psychology along with design characteristics to explain the degree to which users accept and use a computational technology. In their model, a user's attitude toward using a system (affective response) after encountering its design features (external stimulus) is mediated through their perceptions of the system's usefulness and ease of use (cognitive response) to determine their actual system use (behavioral response).

The TAM in turn builds on a psychological framework originated by Fishbein & Azjen, the *Theory of Reasoned Action* [37] (TRA). The basic theory posits that behavior is preceded by intention, with intention in turn determined by an individual's attitude toward the behavior (positive or negative) along with subjective norms, e.g. whether the behavior is seen as appropriate in context or socially acceptable. Azjen's related *Theory of Planned Behavior* [3] added a third determinant of behavior intention, the individual's perception of behavioral control; he also noted the importance of *actual* behavioral control in moderating intention and perceived control, as no one can act if they are not able to do. Venkatesh incorporated these factors in his

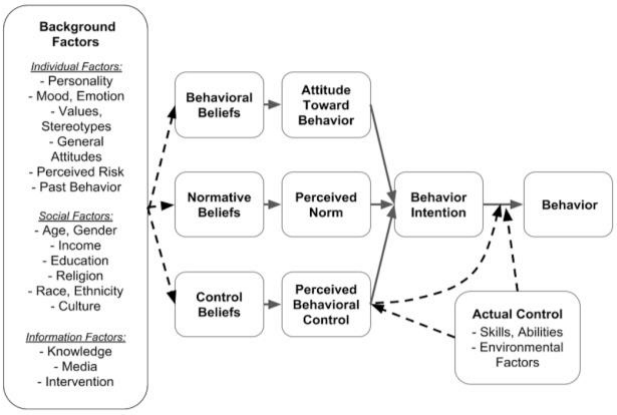



**Figure 1: The Theory of Reasoned Action (TRA)**

work with Davis and others to update the TAM as the *Unified Theory of the Acceptance and Use of Technology* [71,72].

Based on this literature on user behaviors, we incorporated measures of other concepts beyond attitude that we theorized would relate with it, such as privacy, self-efficacy and internet know-how, and also individual measures by which SA-6 would be expected to vary, such as past experience of security breaches, age, and socioeconomic status.

### 2.4. Security Behavior Intentions Scale

The current state of the art for quantifying users' thinking about security practices is the Security Behavior Intentions Scale [32,33] (SeBIS). Its asks about intended user behavior in four areas: device securement, updates, password management, and proactive awareness. SeBIS is not worded as a traditional intention survey – instead of "I intend" statements, it measures intention by asking respondents for their frequency from "Never" to "Always" of such active statements as "I use a password/passcode to unlock my laptop or tablet" – but it has been extensively validated [32,64] and cited by other usable security researchers [31,60]. Its short length makes it practical to include in a larger survey or battery of psychological tests, or to administer during a lab experiment.

However, SeBIS is not a measure of attitudes. Its 16 items are not designed to measure a user's beliefs or emotions about the included behaviors, nor to indicate whether they are attuned to social or situational norms around the behaviors. They do not measure the extent to which the user has the requisite awareness, perceived ability or relevant knowledge to perform the behavior. We see a need for a complementary self-report measure that more directly gets at people's security attitudes underlying their intentions and behavior. We also see a need for a measure that is not tied to technology-specific language, so that the measure retains validity as the security technology changes.

## 3. Consideration of Broader Impacts [77]

We believe a new psychometric scale for assessing a person's attitudes toward expert-recommended tools and practices will be a net benefit to the usable security field and to humanity. While the potential for abuse of such technologies has become recently prominent [40], we have also noted the significant impact to world events from human lapses in cybersecurity judgment [55,79]. SA-6 can help researchers and practitioners to design products in good faith that strengthen resilience to attacks.

## 4. Scale Development and Testing

In developing our self-report attitude measure, we relied on the guidance provided by sources such as Fowler [39], Hinkin et al. [46], Netemeyer et al. [53] and Dillman et al. [28], as well as our own experience and that of our colleagues. Briefly, we sought to measure whether the scale is reliable and valid through analyses of the measures in the literature that we identified that fit with the Theory of Reasoned Action and with security sensitivity literature. We look at the convergence of our scale with these related scales and how the scale varies according to how related measures vary. We iterated in stages to develop a suitable list of candidate items and a survey for testing these items. All pilot work was conducted in accordance with the policies and approval of our Institutional Research Board, as required by U.S. National Science Foundation grant no. CNS-1704087.

### 4.1. Item Generation

A common best practice in psychometric scale development is to generate a long list of possible statements that could measure the underlying construct, in order to increase the chances of developing a sufficiently reliable and valid scale [28,39,46,53]. We generated 200+ items to be rated on a 5-point Likert-type agreement scale (1=*Strongly disagree*, 5=*Strongly agree*). We based the wordings of these items primarily in empirical research by Das et al., but also borrowed some wordings from SeBIS, from other work in usable security and psychology [1,15,16,27,29,41,45,58,66] and from our experiences. We conducted multiple rounds of review of these items, first with experts in usable security who checked the items for content adequacy, then with several nonexperts in security research, whose feedback was used to ensure the survey protocol was clear, unambiguous and easily understandable, in line with common best practices and [28,39,46,53]. These reviews pared our list of items to 60 for online testing.

### 4.2. Survey Development

Another best practice for scale development is to collect variables that are thought to relate with or to vary with the construct, to test if they relate with and vary with the scale to a similar extent [28,39,46,53]. We used the Theory of Reasoned Action [37] as our guide to which constructs we should include measures of in our survey instrument so that we could test for our scale's degree of associations and variances with these constructs. We referred to prior work such as [32,33] for identifying measures for *need for cognition* [14], *consideration of future consequences* [68],

*risk perception* and *risk taking* [11], and *impulsiveness* [67]. We incorporated measures of *internet* and *technical know-how* [48], *computer confidence* [38], and *web-oriented digital literacy* [44], along with general and social *self-efficacy* [66] and the "Big Five" *personality factors* [42]. We included two measures of *privacy concerns* [13,51], a subjective norm strongly related to security beliefs [50,54]. To help test for expected variances in security sensitivity, we asked participants the extent to which *they*, or *someone close to them*, *had been a victim of a security breach*, as well as how much they had *heard or read about security breaches during the past year*. See Section 12.1 for the list of measures included in this report.

Our questionnaire was piloted on Amazon Mechanical Turk with three Masters-qualified workers. Each provided their feedback and suggestions for improving the survey experience via an open-ended text box added at the end. The pilot survey designs ranged between 18 and 24 pages in length as we experimented with how best to break the items among pages and provide clear instructions on each page. The survey was structured to front-load the most important questions, namely the candidate items and the SeBIS questions, because of concern for answers being affected by response fatigue due to the survey length. After the third iteration received entirely positive feedback from a Masters worker, we submitted a formal modification to our Institutional Research Board for review of our survey and research design and exemption from human subjects regulation under U.S. 45 CFR 46.

### 4.3. Finalizing Candidate Items

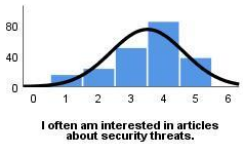
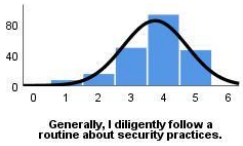
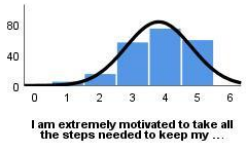
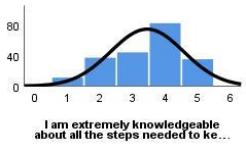
A third best practice for scale development is to collect an initial batch of data to determine which candidate scale items meet minimum standards for response variance and for factor and reliability statistics [28,39,46,53]. We administered 60 candidate items on five pages of 12 items each, along with 16 pages of measures theorized to relate to our survey, in a first round of MTurk research in November-December 2017. We advertised this as a "Survey on attitudes & behaviors among computer users (~30 minutes)" and requested U.S. residents age 18 or older. Using a base rate of $10/hour and a median pilot duration of 24 minutes, we compensated participants with $5 per survey. The survey used one open-ended item asking participants to either mention other security measures they use or to write "None"; this was partly included as a check on attention (if left blank) and fraudulent responses (if nonsensical). As a precaution against workers taking the survey more than once under different IDs, we removed all but one response from an IP address and/or specific location.

We performed an exploratory factor analysis (EFA) and reliability analysis on the $N$ = 196 completed and valid responses. After finding smaller-than-desired variance in some response distributions and in the alpha and total variance explained by items with high factor loadings, we decided to retain just 18 items without changes. We

**Table 1: Sample statistics for scale finalizing, validation**

| | *Scale finalizing* | *Validity study* |
|---|---|---|
| ***N*** | ***478*** | ***209*** |
| ***What is your age range?*** | | |
| 18-29 | 46.7% | 20.6% |
| 30-39 | 32.2% | 18.7% |
| 40-49 | 10.0% | 23.0% |
| 50-59 | 7.1% | 21.5% |
| 60 or older | 4.0% | 16.3% |
| ***What is your gender identity?*** | | |
| Male | 41.6% | 41.6% |
| Female | 57.7% | 58.4% |
| Nonbinary or non-conforming | 0.6% | 0.0% |
| ***What is your level of education?*** | | |
| Some high school | 0.6% | 0.0% |
| High school degree/equivalent | 7.3% | 32.5% |
| Some college/assoc./tech. deg. | 28.7% | 37.8% |
| Bachelor's degree | 43.7% | 15.3% |
| Graduate/professional degree | 19.7% | 14.4% |
| ***Are you a U.S. citizen?*** | | |
| Yes | 91.2% | 98.1% |
| No | 8.8% | 1.9% |
| ***What is your yearly household income?*** | | |
| Up to $25,000 | 24.5% | 22.5% |
| $25,000 to $49,999 | 29.1% | 24.9% |
| $50,000 to $74,999 | 18.8% | 33.5% |
| $75,000 to $99,999 | 13.2% | 9.1% |
| $100,000 or more | 14.4% | 10.0% |
| ***What is your employment status?*** | | |
| Employed full time | (Not Asked) | 42.1% |
| Employed part time | | 8.1% |
| Unemployed looking for work | | 9.6% |
| Unemployed not looking for work | | 8.6% |
| Retired | | 16.7% |
| Student | | 5.7% |
| Disabled | | 9.1% |
| ***How frequently or infrequently have you personally been the victim of a breach of security (e.g. hacking, viruses, theft of personal data)?*** | | |
| Very infrequently | 51.3% | 39.7% |
| Infrequently | 27.0% | 29.2% |
| Neither infrequently or frequently | 11.7% | 18.2% |
| Frequently | 8.2% | 10.0% |
| Very frequently | 1.9% | 2.9% |
| ***How frequently or infrequently has someone close to you (e.g. spouse, family member or close friend) been the victim of a breach of security (e.g. hacking, viruses, theft of personal data)?*** | | |
| Very infrequently | 28.2% | 28.7% |
| Infrequently | 40.6% | 34.4% |
| Neither infrequently or frequently | 18.4% | 23.9% |
| Frequently | 11.7% | 9.6% |
| Very frequently | 1.0% | 3.3% |
| ***How much have you heard or read during the last year about online security breaches?*** | | |
| None at all | 5.0% | 7.2% |
| A little | 32.4% | 24.9% |
| A moderate amount | 35.4% | 38.8% |
| A lot | 19.0% | 21.1% |
| A great deal | 8.2% | 8.1% |

**Table 2: Final set of SA-6 items with factor loadings, alpha if item deleted, and histograms. Factor loadings well above .40 indicate strong relationships. Alpha above .70 indicates strong internal consistency of scale responses.**

| *SA-6 scale items* (Principal Components Analysis; Overall alpha: .84) | *Factor loading* | *Alpha if item deleted* | *Histograms* (1=Strongly disagree, 5=Strongly agree) |
|---|---|---|---|
| **I seek out opportunities to learn about security measures that are relevant to me.** | 0.81 | 0.80 |   |
| **I am extremely motivated to take all the steps needed to keep my online data and accounts safe.** | 0.78 | 0.81 |   |
| **Generally, I diligently follow a routine about security practices.** | 0.77 | 0.81 |   |
| **I often am interested in articles about security threats.** | 0.72 | 0.82 |   |
| **I always pay attention to experts' advice about the steps I need to take to keep my online data and accounts safe.** | 0.71 | 0.83 |   |
| **I am extremely knowledgeable about all the steps needed to keep my online data and accounts safe.** | 0.71 | 0.83 |   |

generated 30 new items that used more extreme wordings to encourage a greater response distribution. After reviews similar to those in Section 4.1, a final list of 48 candidate items were deployed in an MTurk survey in February 2018; on these newly gathered $N$ =339 responses, we performed several EFAs and reliability analyses and examined the item response distributions, factor loadings, factor alphas, and total variances explained to ensure that they displayed sufficient psychometric properties for further testing. The 48 candidate items and their sources are listed in Section 12.2.

### 4.4. Finalizing Scale Items

The next stage of scale development was to collect a sufficient number of responses from which to narrow the list of items to those that most clearly measured the security sensitivity construct [28,39,46,53]. To this end, in July-August 2018, we collected a third dataset on MTurk and a fourth dataset in a university-run online study pool, using very similar recruitment language and the same participant compensation as in Section 4.3. A chi-square analysis found that these datasets did not differ significantly by gender: $X^2(1, N = 475) = 2.95$, $p$ = n.s. We conducted 10 pairwise comparisons of the datasets by age range, first correcting for possible compounded Type I error by conducting a Bonferroni procedure that adjusted alpha to $p < .005$. We did not find any pairwise comparisons by age to be statistically significant: overall $X^2(4, N=478)=11.42$, $p$ = n.s. See Section 12.3 for chi-square statistics for age-level pairwise comparisons and for the pairwise comparisons by levels of education, income and breach-experience measures.

Based on the lack of significant differences by age or gender, we merged these to form one sample of $N$=479. This ensured a 5:1-to-10:1 ratio of observations to items for finalizing the scale, as recommended by [39,46,53]. See Table 1 for descriptive statistics for this sample.

We conducted a series of factor analyses and reliability assessment to identify the number of possible factors from scree plots and factor loadings; which factors explained at least 40% total variance in the sample and eigenvalues over 1.0; and which of these factors met a threshold Cronbach's alpha of .70. Finally, we tested the goodness-of-fit of each candidate factor structure by conducting a confirmatory factor analysis (CFA) to calculate fit statistics that are appropriate for a large sample [47]: the Comparative Fit Index (CFI), for which an acceptable fit is above .90 and a superior fit above .95, and the Standardized Root Mean Square Residual (SRMR), which should be below .08. We chose the first factor, which explained 64% of the sample variance with 6 items loading over 0.71 on this factor. These six items had a Cronbach's alpha equal to .88, demonstrating excellent internal reliability (well above the threshold of .70); and a CFI of 0.96 and SRMR of 0.03, demonstrating superior model fit. Section 12.4 displays the six item histograms, factor loadings and alpha if item deleted.

## 5. Validity Study in Census-tailored Panel

To test the reliability and validity of SA-6 outside of the MTurk and university study populations, we repeated our study in September 2018 with a U.S. Census-tailored panel filled by Qualtrics ($N$=209). We again targeted compensation at $5 per response, however this was not handled by us directly; Qualtrics worked with its third-party providers to provide sufficient payment in forms such as reward points.

We dropped survey measures that were less central to this report, reordered items so that the demographics questions were asked first to fill the survey quotas, added a question about employment status, and (beyond the open-ended item noted in Section 4.3) added a second attention check: "*We use this question to discard the answers of people who are not reading the questions. Please select "51% to 75% of the*

*time" (option 4) to preserve your answers*." The panel received a sufficient number of responses in all variable categories to complete the statistical picture for this report. See Table 1 for descriptive statistics for this sample.

As before, we examined the items' statistical properties and confirmed the factor structure in this smaller sample. SA-6 was found to explain 56% of total sample variance, with a Cronbach's alpha of .84, a CFI of 0.91, and an SRMR of 0.05. Table 2 displays the six item histograms, factor loadings and statistics for Cronbach's alpha if item deleted for SA-6. These demonstrate SA-6's solid factor structure, internal consistency and goodness of fit.

## 6. Convergent and Discriminant Validity

In the Census-tailored sample ($n$=209), we conducted a series of correlations and independent-samples t-tests to assess the degree to which security attitudes as measured by SA-6 converged with measures thought to relate with it (*convergent validity*) and varied as expected by categorical measures (*discriminant validity*), consistent with the Theory of Reasoned Action [37]. These tests support that the scale is measuring the concept that it claims to measure. We excluded some collected variables from validity tests because they did not meet a Cronbach's alpha of .70, which indicates they may include higher-than-acceptable random measurement error. Section 12.5 reports the Cronbach's alpha values for each observed measure.

### 6.1. Correlation with SeBIS

To examine convergent validity of SA-6, we first tested its statistical association with SeBIS, the field's standard self-report measure of security behavior intention. We did this because attitude is a direct antecedent of behavior intention in the Theory of Reasoned Action [37]. Using a Spearman correlation, we found SA-6 to be significantly positively associated with SeBIS ($r$ =.54, $p$ < .01). Using linear regression, we found that SA-6 explained 28% of the variance in SeBIS ($p$<.01). This result is consistent with longstanding psychological evidence of the relationship between attitudes and behavior intention [5,7,8,37,69,75] and demonstrates SA-6's convergent validity.

### 6.2. Correlations with Other Interval Variables

To further examine the convergent validity of SA-6, we looked at its statistical association with measures of perceived behavioral control, perceived norms (chiefly privacy) and individual cognitive and risk styles. We collected and tested these measures because these were used in validity testing for SeBIS [32–34] since they represent closely associated concepts. These concepts are also components of the Theory of Reasoned Action [37].

We found expected significant associations among SA-6 and psychological indicators of perceived behavioral control (Barratt Impulsiveness Scale $r$= -.180, $p$<.01; General Self-Efficacy, $r$=.208, $p$<.01; Social Self-Efficacy, $r$=.363, $p$<.01); indicators of privacy concerns (Internet Users' Informational Privacy Concerns $r$=.390, $p$<.01; Privacy Concerns Scale ($r$=.382, $p$<.01); and two indicators of cognition and risk styles (Need for Cognition $r$=.258, $p$<.01, and the Domain-Specific Risk Taking Health/Safety subscale for risk perception: $r$=.175, $p$<.05). We did not find a significant association for SA-6 with the Consideration of Future Consequences scale, with the General Decision-Making Styles subscales for dependence and avoidance, or with the Domain-Specific Risk-Taking Health/Safety subscale for risk-taking propensity.

We found a significant association of SA-6 with the "Big Five" personality factor of Extraversion ($r$=.175, $p$<.05). We included the Big 5 because personality is a background component of the Theory of Reasoned Action [37].

We found an expected significant positive correlation with the Kang Internet Know-How scale ($r$=.542, $p$<.01) and with two related scales, one for confidence in using computers ($r$=.280, $p$<.01) and the other for web-oriented digital literacy ($r$=.503, $p$<.01). We included these measures because information, skill and ability are key components of the Theory of Reasoned Action [37].

### 6.3. Variances by Categorical or Ordinal Variables

To examine discriminant validity, we tested whether SA-6 varied significantly as a function of personal experiences of and media exposure to security breaches, and by age, gender and socioeconomic status. We included these measures because social and informational measures are antecedents of attitude in the Theory of Reasoned Action [37] and previous work has found a connection between demographics and security concern [50,54].

For each type of experience with security breaches, we recoded the 5-level variable responses into 2 levels (low experience (1-2) vs. high experience (3-5)) and conducted independent-samples $t$-tests on the census-weighted sample. This analysis let us look at how SA-6 varied for people with low versus high levels of experience with security breaches (see Table 3 for a summary). SA-6 was significantly higher for participants with higher self-reported frequency of participants falling victim to a security breach, higher self-reported frequency of their close friends or relatives falling victim. and by the amount they had heard or seen about security breaches in the past year.

For demographics, we found a statistically significant difference in SA-6 by age group and gender, with a higher score for older participants and men. SA-6 scores were also higher for participants who attended college and those whose yearly household income exceeded the 2018 U.S. poverty level of $25,100 for a family of four [80]. These differences correspond with differences observed in other studies on cybersecurity opinions and knowledge [50,54]. We did not find a significant difference in SA-6 by citizenship or employment status, with the exception of "Employed full-time" ($M$=3.85, $SD$=.75) vs. "Unemployed looking for work" ($M$=3.24, $SD$=.76, $F$(6,202)=2.59, $p$<.05).

**Table 3: SA-6 Mean, standard deviation, and test of difference for security breach experience and demographic variables**

| | SA-6 Mean (SD) | | t(df), p |
|---|---|---|---|
| **Security breach experience frequency** | | | |
| | ***Low*** | ***High*** | |
| Themselves falling victim to a security breach | 3.56 (.78) | 4.13 (.58) | $t(41.46) = -4.54$, $p<.001$ |
| Close friends or relatives falling victim to a breach | 3.57 (.76) | 4.10 (.74) | $t(207)= -3.40$, $p<.005$ |
| Heard about security breaches in the past year | 3.35 (.80) | 3.77 (.74) | $t(207)=-3.77$, $p<.001$). |
| **Demographic differences** | | | |
| | ***18-39*** | ***40 +*** | |
| Age group | 3.40 (.81) | 3.69 (.76) | $t(207)= -2.172$, $p<.05$ |
| | ***Male*** | ***Female*** | |
| Gender | 3.77 (.71) | 3.53 (.81) | $t(198.38)= 2.19$, $p<.05$ |
| | ***No college*** | ***Attend. college*** | |
| College attendance | 3.42 (.79) | 3.73 (.76) | $t(207)=-2.76$, $p<.01$ |
| | ***Below $25K*** | ***Above $25K*** | |
| Income level | 3.30 (.71) | 3.73 (.77) | $t(207)=-3.42$, $p<.005$ |

### 6.4. Variances by Participants' Recall of Security Actions

We were able to go one step further than the authors of SeBIS and ask respondents whether, in the past week, they had at least once taken an expert-recommended action for device securement, updating, password management or proactive awareness. The item wordings were drawn from those of SeBIS in those areas, with a response set of "Yes/No/Not Sure/NA" and these instructions: "*For the following statements, please select the response that best represents your recall of what actions you have taken in the past week. Please select "I'm not sure" if you don't know the answer. Please select "NA" if the statement does not apply to you*." We excluded NA responses from the item-level analysis. We recoded the remaining 3-level variable responses into 2 levels (Yes (1) vs. No or Not Sure (2-3)) and conducted independent-samples *t*-tests on the census-weighted sample. This analysis let us look at how SA-6 varied for people who did vs. did not recall performing these certain SeBIS-derived security actions. We found SA-6 to vary significantly by the answers to all but one item. This further demonstrates discriminant validity. See Table 4 for item statistics.

We then conducted a series of binary logistic regressions to compare predicted outcomes by (a) models that combined SA-6 with SeBIS as predictors, (b) models using SeBIS without SA-6 as a predictor, and (c) models using only a constant as a predictor (to indicate baseline performance without SeBIS). Results indicated that there was a significant association among SA-6, SeBIS and item responses. This improved the performance of models for three items: "*In the past week, I have downloaded and installed at least one available update for my computer's operating system within 24 hours of receiving a notification that it was available"* ($X^2(2) = 42.49$, $p < .001$), boosting the model's percentage of correctly classified responses to (a) 68.1% vs. (b) 67.5% for SeBIS without SA-6 and (c) 58.6% for the constant alone; "*In the past week, I have verified at least once that I am running antivirus software that is fully updated"* ($X^2(2) = 43.06$, $p < .001$), boosting the model's percentage of correctly classified responses to (a) 65.9% vs. (b) 64.9% for SeBIS without SA-6 and (c) 52.7% for the constant alone; and "*In the past week, I have used a password/passcode at least once to unlock my tablet"* ($X^2(2) = 39.65$, $p < .001$), boosting the model's percentage of correctly classified responses to (a) 77.2% vs. (b) 76.7% for SeBIS without SA-6 and (c) 70.5% for the constant alone. See Section 12.5 for the classification tables for each item's logistic regressions.

The pseudo R-squared value generated with logistic regressions cannot be said, as with a linear regression R-squared value, to show the variance accounted for by the model. In order to use a linear regression model to calculate this variance explained, we transformed the recalled security action items into one interval variable by computing an average of the scores of the nine items that were found to vary significantly by their SA-6 score. The Cronbach's alpha for this compound measure was .77, comfortably above the threshold of .70 we used to exclude measures from validity tests. When we combined SA-6 with SeBIS as a predictor in this model, SA-6 lifted its ability to explain the variance in this compound measure from 23.5% to 24.3% ($p<.001$, $r=.493$). A Spearman correlation also found significant associations (SA-6 with recalled security actions: $r=.398$, $p<.001$; SeBIS with recalled security actions: $r=.541$, $p<.001$). These statistics suggest that SeBIS is a partial mediator of SA-6's influence on the recalled security actions measure, as predicted by the Theory of Reasoned Action's model of attitude helping to determine behavior through the mediation of intention [37].

## 7. Discussion and Future Work

Our careful scale development process gives us confidence that SA-6 has demonstrated construct validity, internal consistency and reliability, goodness-of-fit, and convergent and discriminative validity. We conducted several tests of our generated items to determine which were most suitable for our scale, then visually inspected the response distributions and conducted factor and reliability analyses to determine which mix of items are the best fit for a short self-report measure of security attitudes. We found SA-6 to correlate as expected with privacy and other theorized concepts such as self-efficacy, and to vary by factors such as exposure to breaches and demographics.

**Table 4: Means, standard deviations, and T statistics for participants' answers to recalled security action statements**

| | Yes | | No or Not Sure | | | | NA |
|---|---|---|---|---|---|---|---|
| *Participant's recalled security action* | *M* | *SD* | *M* | *SD* | *t* | *df* | *%* |
| In the past week, I have changed a password for at least one of my online accounts. | 3.84 | 0.77 | 3.39 | 0.74 | 4.20**** | 200 | 3.3 |
| In the past week, I have downloaded and installed at least one available update for my computer's operating system within 24 hours of receiving a notification that it was available. | 3.95 | 0.73 | 3.40 | 0.75 | 5.11**** | 189 | 8.6 |
| In the past week, I have left my laptop or desktop computer unlocked at least once when I walked away from it. | 3.49 | 0.81 | 3.84 | 0.70 | 2.95*** | 184 | 11.0 |
| In the past week, I have submitted information to a website at least once without first verifying that it would be sent securely. | 3.64 | 0.76 | 3.68 | 0.78 | (n.s.) | 194 | 6.2 |
| In the past week, I have used a password/passcode at least once to unlock my tablet. | 3.71 | 0.80 | 3.40 | 0.68 | 2.56* | 191 | 7.7 |
| In the past week, I have used at least one password that contains 10 or more characters. | 3.77 | 0.75 | 3.39 | 0.77 | 3.43*** | 203 | 1.9 |
| In the past week, I have used the exact same password for at least two online accounts. | 3.49 | 0.82 | 3.82 | 0.69 | 2.95*** | 200 | 3.3 |
| In the past week, I have verified at least once that I am running antivirus software that is fully updated. | 3.82 | 0.69 | 3.49 | 0.82 | 3.03*** | 184 | 1.9 |
| In the past week, I have verified that at least one app or software program that I use is fully updated. | 3.80 | 0.76 | 3.38 | 0.75 | 3.84**** | 200 | 3.3 |
| In the past week, I have verified the URL of at least one internet link that I received in email before deciding whether to click on it. | 3.94 | 0.64 | 3.45 | 0.77 | 4.71**** | 189 | 8.6 |

*Sig. at .05 level ** Sig. at .01 level ***Sig. at .005 level ****Sig. at .001 level

### 7.1. Using SA-6 to Measure Security Attitudes

SA-6 will be useful to researchers and practitioners who need a reliable and valid method to systematically assess and compare user attitudes about the use and adoption of expert-recommended security tools and practices. SA-6 is easily administered via an online questionnaire in a web browser or on paper, and it also is shorter than measures such as the 31-item Personal Data Attitude measure for adaptive cybersecurity [2] or 63-item Human Aspects of Information Security Questionnaire [56]. SA-6 and its individual subscales will help to answer research questions such as: *To what degree does a user report the awareness, motivation or knowledge to perform recommended security actions? How positive or negative is her or his attitude? To what degree is she or he likely to consider adopting more-secure tools? How does her or his score compare with a group average?*

SA-6's usefulness is not constrained to research motivated by the Theory of Reasoned Action/Theory of Planned Behavior or the Technology Acceptance Models. Our measure could contribute a valuable tool to research motivated by *Self-Determination Theory* [63], helping to assess intrinsic motivation to use recommended security tools and practices. It also could be used for approximating threat appraisal in adaptions to usable security of *Protection Motivation Theory* [52,62] and for pre- and post-study evaluations in cybersecurity education research [19].

### 7.2. Using SA-6 to Predict Security Behaviors

An open question in psychology is the degree to which attitude, intention and other factors directly determine behavior. Sutton's 1998 meta-analyses [69] showed that TRA and TPB explain on average between 40% and 50% of intention variance, with the rest accounted for by changes in factors such as volitional control and random variance. And Webb and Sheeran's 2006 meta-analyses [75] showed that across 47 experiments, a medium-to-large change in intention ($d$=0.66) led to a small-to-medium change in behavior ($d$=0.36). They conclude that "intentional control of behavior is a great deal more limited than previous meta-analyses of correlational studies have indicated."

Sutton notes a relevant distinction in this context between explanation and prediction. In his framing, explanation is a process of identifying what determines intentions and behavior and seeking how such factors combine, while prediction enables the targeting of interventions in spite of not understanding the full degree and nature of a behavior's determinants. An example Sutton gives of the latter is identification of people at high risk of developing a drinking problem, arguing that, despite not having a clear model of which factors combine to influence alcohol addiction, it is still a benefit to create predictive models of alcoholism risk in order to target an early intervention. Nevertheless, he writes, a causal model that sheds light on what factors influence drinking in certain individuals may make it possible to extend the predictive model to similar problems and to avoid a "one size fits all" solution that can better target interventions by differing nature and content.

Similarly, our results suggest to us that even given the moderate $r$ values shown in our correlation analyses, SA-6 is likely to add valuable predictive weight with SeBIS in computational modeling of who is likely to act on security recommendations and who is open to changing their security behavior. We see both scales as useful for future research into the degree to which security sensitivity along with security behavior intention can explain which architecture choices or "nudges", as suggested by Egelman & Peer

[34,35], and Redmiles et al. [59] might best improve security choices by users with specific attitude and intention profiles, thus helping the field move beyond a blanket approach to interventions. We also are pursuing work to compare SA-6 with two other scales we are developing to measure users' concernedness with and resistance to changing their security behaviors, as part of a new causal model and framework.

#### 7.3. Using SA-6 vs. SeBIS to Identify Target Interventions

We see utility for SA-6 in measuring a user's readiness for educational interventions. Broadly, SA-6 identifies whether the user is a good candidate for interventions of two types: (1) to raise awareness of the general need for using expert-recommended tools and practices (*low SA-6*) or (2) to add to users' knowledge of how to use recommended tools and practices (*high SA-6*). An example of the first type of intervention might be playing a security awareness game, while an example of the second type would be taking part in a tutorial on creating strong but memorable passwords.

Conversely, we see SeBIS as offering specific utility in measuring a user's readiness for motivational interventions that (3) move them into the intention stage (*low SeBIS*) or (4) move them from intention into action and reinforce action (*high SeBIS*). An example of this third type of intervention would be a positive incentive program, such as rewards for 3, 15 and 30 days of consecutive use of a third-party password manager. An example of the fourth type would be reminders to act, such as context-aware notifications of a newly available software update, or negative incentives for nonaction, such as progressively annoying or persistent notifications for a software update that a user fails to install.

## 8. Limitations and Next Steps

Our project was conducted with U.S.-based populations age 18 or older using a lengthy, English-language, online questionnaire. More research will be needed to find support for SA-6's reliability and validity in populations of computer users outside the U.S. and/or when translated into other languages. The ability to generalize our results inside the U.S. is limited by our use of purposive, nonrandom sampling of the subpopulation of online survey-takers. Our use of online surveys as the only method of questionnaire administration may also have introduced common method bias, suggesting the need also to test the survey in other modes such as written and telephone versions.

All correlational research is inherently unable to prove causation. This work is only the first step toward finding support for a relationship among the variables in our study. Experimental research will be needed to investigate the hypothesis that changes in security sensitivity will lead to changes in security behavior intention and, ultimately, to changes in actual security behavior by end users. Additionally, we did not test for measurement noninvariance, which limits SA-6's usefulness for comparing groups. Finally, some items in SeBIS and in the SeBIS-derived items in Table 4 are out of step with current security recommendations (*e.g.,* many experts now advise against forcing users to periodically change their passwords) and features in consumer systems *(e.g.,* many updates can now be downloaded and installed automatically). This limits the usefulness of SeBIS and these SeBIS-derived items for accurately measuring security intention and recalled actions.

However, we believe that SA-6 is a valid way to measure security attitudes for future studies and experiments relevant to cybersecurity. We are pursuing a second work that will allow for a side-by-side discussion of this scale with two others in development and provide support for a causal model and framework for targeting security interventions.

## 9. Conclusion

In this paper, we introduce and validate SA-6, a self-report measure of end-user security attitudes. Using principles of psychological scale development, we generated and finalized six items that (a) correspond to prior work on their face; that (b) pilot testers found to be unambiguous and easily answered; that (c) demonstrated sufficient response variance, and that (d) were found in factor and reliability analyses to demonstrate desired psychometric properties.

Via analyses of data from a U.S. Census-tailored survey panel, we found SA-6 to be significantly associated with a self-report measure of behavior intention and to exhibit expected variances by participants' recollections of recent security actions. We found SA-6 significantly associated with other measures of cognition and with measures of subjective norms, chiefly privacy, and perceived behavioral control, such as self-efficacy and internet know-how.

Our scale is a lightweight tool for researchers and practitioners to (1) quantify and compare end users' attitudes toward using recommended security tools and practices, and (2) improve predictive modeling of who will adopt such behaviors. The field of usable security will benefit from this systematic method for assessing a user's awareness, motivation, and knowledge of expert-recommended tools and practices. We hope our work helps improve understanding of end-user compliance with security recommendations and the identification of users who are susceptible to attacks and open to changing their behaviors.

## 10. Acknowledgments

This research was sponsored by the U.S. National Science Foundation under grant no. CNS-1704087. We thank Sauvik Das for his feedback on early versions of this scale, Maria Tomprou for her advice about statistical analysis and framing, and Geoff Kaufman and members of the CoEx Lab and CHIMPS Lab at CMU for helping us to think through the many iterations of this research and analysis.

We also thank our CHI and SOUPS reviewers for their thoughtful critiques, which improved this research paper.

## 11. References


[1] Anne Adams and Martina Angela Sasse. 1999. Users are not the enemy. *Commun. ACM* 42, 12 (December 1999), 40–46. DOI:https://doi.org/10.1145/322796.322806

[2] Joyce Hoese Addae, Michael Brown, Xu Sun, Dave Towey, and Milena Radenkovic. 2017. Measuring attitude towards personal data for adaptive cybersecurity. *Inf. Comput. Secur.* 25, 5 (October 2017), 560–579. DOI:https://doi.org/10.1108/ICS-11-2016-0085

[3] Icek Ajzen. 1991. The theory of planned behavior. *Organ. Behav. Hum. Decis. Process.* 50, 2 (December 1991), 179–211. DOI:https://doi.org/10.1016/0749-5978(91)90020-T

[4] Icek Ajzen. 2001. Nature and Operation of Attitudes. *Annu. Rev. Psychol.* 52, 1 (2001), 27–58. DOI:https://doi.org/10.1146/annurev.psych.52.1.27

[5} Icek Ajzen, 2006. Behavioral Interventions Based on the Theory of Planned Behavior.

[6] Icek Ajzen and Martin Fishbein. 2000. Attitudes and the Attitude-Behavior Relation: Reasoned and Automatic Processes. *Eur. Rev. Soc. Psychol.* 11, 1 (January 2000), 1–33. DOI:https://doi.org/10.1080/14792779943000116

[8] Icek Ajzen and Thomas J Madden. 1986. Prediction of goal-directed behavior: Attitudes, intentions, and perceived behavioral control. *J. Exp. Soc. Psychol.* 22, 5 (September 1986), 453–474. DOI:https://doi.org/10.1016/0022-1031(86)90045-4

[9] Sebastian Bamberg. 2003. How does environmental concern influence specific environmentally related behaviors? A new answer to an old question. *J. Environ. Psychol.* 23, 1 (March 2003), 21–32. DOI:https://doi.org/10.1016/S0272-4944(02)00078-6

[11] Ann-Renee Blais and Elke Weber. 2006. *A Domain-Specific Risk-Taking (DOSPERT) Scale for Adult Populations*. Social Science Research Network, Rochester, NY. Retrieved November 14, 2017 from https://papers.ssrn.com/abstract=1301089

[12] C. Bravo-Lillo, L. F. Cranor, J. Downs, and S. Komanduri. 2011. Bridging the Gap in Computer Security Warnings: A Mental Model Approach. *IEEE Secur. Priv.* 9, 2 (March 2011), 18–26. DOI:https://doi.org/10.1109/MSP.2010.198

[13] Tom Buchanan, Carina Paine, Adam N. Joinson, and Ulf-Dietrich Reips. 2007. Development of measures of online privacy concern and protection for use on the Internet. *J. Am. Soc. Inf. Sci. Technol.* 58, 2 (January 2007), 157–165. DOI:https://doi.org/10.1002/asi.20459

[14] John T. Cacioppo, Richard E. Petty, and Chuan Feng Kao. 1984. The Efficient Assessment of Need for Cognition. *J. Pers. Assess.* 48, 3 (June 1984), 306–307. DOI:https://doi.org/10.1207/s15327752jpa4803_13

[15] Robert B. Cialdini. 2001. *Influence: science and practice* (4th ed ed.). Allyn and Bacon, Boston, MA.

[16] Robert B. Cialdini and Noah J. Goldstein. 2004. Social Influence: Compliance and Conformity. *Annu. Rev. Psychol.* 55, 1 (January 2004), 591–621. DOI:https://doi.org/10.1146/annurev.psych.55.090902.142015

[17] Lorrie Faith Cranor. 2008. A Framework for Reasoning About the Human in the Loop. In *Proceedings of the 1st Conference on Usability, Psychology, and Security* (UPSEC'08), 1:1–1:15. Retrieved September 21, 2018 from http://dl.acm.org/citation.cfm?id=1387649.1387650

[18] Jonas Dalege, Denny Borsboom, Frenk van Harreveld, Helma van den Berg, Mark Conner, and Han L. J. van der Maas. 2016. Toward a formalized account of attitudes: The Causal Attitude Network (CAN) model. *Psychol. Rev.* 123, 1 (2016), 2–22. DOI:https://doi.org/10.1037/a0039802

[19] M. Dark and J. Mirkovic. 2015. Evaluation Theory and Practice Applied to Cybersecurity Education. *IEEE Secur. Priv.* 13, 2 (March 2015), 75–80. DOI:https://doi.org/10.1109/MSP.2015.27

[20] Sauvik Das. 2017. Social Cybersecurity: Reshaping Security Through An Empirical Understanding of Human Social Behavior. *Dissertations* (May 2017). Retrieved from http://repository.cmu.edu/dissertations/982

[21] Sauvik Das, Tiffany Hyun-Jin Kim, Laura A. Dabbish, and Jason I. Hong. 2014. The effect of social influence on security sensitivity. In *Proc. SOUPS*. Retrieved from https://pdfs.semanticscholar.org/cd64/4bceb458cfb63c16a86fdf0234c8cf54c004.pdf

[22] Sauvik Das, Adam D.I. Kramer, Laura A. Dabbish, and Jason I. Hong. 2014. Increasing Security Sensitivity With Social Proof: A Large-Scale Experimental Confirmation. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security* (CCS '14), 739–749. DOI:https://doi.org/10.1145/2660267.2660271

[23] Sauvik Das, Adam D.I. Kramer, Laura A. Dabbish, and Jason I. Hong. 2015. The Role of Social Influence in Security Feature Adoption. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (CSCW '15), 1416–1426. DOI:https://doi.org/10.1145/2675133.2675225

[24] Sauvik Das, Joanne Lo, Laura Dabbish, and Jason I. Hong. 2018. Breaking! A Typology of Security and Privacy News and How It's Shared. *ACM CHI 2018 Conf. Hum. Factors Comput. Syst.* 1, 1 (2018), 2.

[25] Fred D. Davis. 1989. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Q.* 13, 3 (1989), 319–340. DOI:https://doi.org/10.2307/249008

[26] Fred D. Davis, Richard P. Bagozzi, and Paul R. Warshaw. 1989. User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Manag. Sci.* 35, 8 (August 1989), 982–1003. DOI:https://doi.org/10.1287/mnsc.35.8.982

[27] Carlo C. DiClemente, James O. Prochaska, and Michael Gibertini. 1985. Self-efficacy and the stages of self-change of smoking. *Cogn. Ther. Res.* 9, 2 (April 1985), 181–200. DOI:https://doi.org/10.1007/BF01204849

[28] Don A. Dillman, Jolene D. Smyth, and Leah Melani Christian. 2014. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. John Wiley & Sons.

[29] Paul Dourish, Rebecca E. Grinter, Jessica Delgado de la Flor, and Melissa Joseph. 2004. Security in the wild: user strategies for managing security as an everyday, practical problem. *Pers. Ubiquitous Comput.* 8, 6 (November 2004), 391–401. DOI:https://doi.org/10.1007/s00779-004-0308-5

[30] Alice H. Eagly and Shelly Chaiken. 1993. *The psychology of attitudes*. Harcourt Brace Jovanovich College Publishers, Orlando, FL, US.

[31] Serge Egelman, Lorrie Faith Cranor, and Jason Hong. 2008. You'Ve Been Warned: An Empirical Study of the Effectiveness of Web Browser Phishing Warnings. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '08), 1065–1074. DOI:https://doi.org/10.1145/1357054.1357219

[32] Serge Egelman, Marian Harbach, and Eyal Peer. 2016. Behavior Ever Follows Intention?: A Validation of the Security Behavior Intentions Scale (SeBIS). In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16), 5257–5261. DOI:https://doi.org/10.1145/2858036.2858265

[33] Serge Egelman and Eyal Peer. 2015. Scaling the Security Wall: Developing a Security Behavior Intentions Scale (SeBIS). In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15), 2873–2882. DOI:https://doi.org/10.1145/2702123.2702249

[34] Serge Egelman and Eyal Peer. 2015. Predicting privacy and security attitudes. *ACM SIGCAS Comput. Soc.* 45, 1 (2015), 22–28.

[35] Serge Egelman and Eyal Peer. 2015. The Myth of the Average User: Improving Privacy and Security Systems through Individualization. In *Proceedings of the New Security Paradigms Workshop on ZZZ - NSPW '15*, 16–28. DOI:https://doi.org/10.1145/2841113.2841115

[36] Michael Fagan, Mohammad Maifi Hasan Khan, and Ross Buck. 2015. A Study of Users' Experiences and Beliefs About Software Update Messages. *Comput Hum Behav* 51, PA (October 2015), 504–519. DOI:https://doi.org/10.1016/j.chb.2015.04.075

[37] Martin Fishbein and Icek Ajzen. 2010. *Predicting and changing behavior: The reasoned action approach*. Psychology Press, New York, NY, US.

[38] Gerry Fogarty, Patricia Cretchley, Chris Harman, Nerida Ellerton, and Nissam Konki. 2001. Validation of a questionnaire to measure mathematics confidence, computer confidence, and attitudes towards the use of technology for learning mathematics. *Math. Educ. Res. J.* 13, 2 (2001), 154–160.

[39] Floyd J. Fowler. 1995. *Improving Survey Questions: Design and Evaluation*. SAGE.

[40] McKenzie Funk. 2016. Opinion | Cambridge Analytica and the Secret Agenda of a Facebook Quiz. *The New York Times*. Retrieved March 19, 2018 from https://www.nytimes.com/2016/11/20/opinion/cambridge-analytica-facebook-quiz.html

[41] Shirley Gaw, Edward W Felten, and Patricia Fernandez-Kelly. 2006. Secrecy, Flagging, and Paranoia: Adoption Criteria in Encrypted E-Mail. (2006), 10.

[42] Samuel D Gosling, Peter J Rentfrow, and William B Swann. 2003. A very brief measure of the Big-Five personality domains. *J. Res. Personal.* 37, 6 (December 2003), 504–528. DOI:https://doi.org/10.1016/S0092-6566(03)00046-1

[43] Heesup Han and Hae Jin Yoon. 2015. Hotel customers' environmentally responsible behavioral intention: Impact of key constructs on decision in green consumerism. *Int. J. Hosp. Manag.* 45, (February 2015), 22–33. DOI:https://doi.org/10.1016/j.ijhm.2014.11.004

[44] Eszter Hargittai. 2005. Survey Measures of Web-Oriented Digital Literacy. *Soc. Sci. Comput. Rev.* 23, 3 (August 2005), 371–379. DOI:https://doi.org/10.1177/0894439305275911

[45] Cormac Herley. 2009. So Long, and No Thanks for the Externalities: The Rational Rejection of Security Advice by Users. In *Proceedings of the 2009 Workshop on New Security Paradigms Workshop* (NSPW '09), 133–144. DOI:https://doi.org/10.1145/1719030.1719050

[46] Timothy R. Hinkin, J. Bruce Tracey, and Cathy A. Enz. 1997. Scale Construction: Developing Reliable and Valid Measurement Instruments. *J. Hosp. Tour. Res.* 21, 1 (February 1997), 100–120. DOI:https://doi.org/10.1177/109634809702100108

[47] Daire Hooper, Joseph Coughlan, and Michael Mullen. Structural Equation Modelling: Guidelines for Determining Model Fit. 11.

[38] Ruogu Kang, Laura Dabbish, Nathaniel Fruchter, and Sara Kiesler. 2015. "My data just goes everywhere:" User mental models of the internet and implications for privacy and security. In *Symposium on Usable Privacy and Security (SOUPS)*, 39–52.

[49] Stephen J. Kraus. 1995. Attitudes and the Prediction of Behavior: A Meta-Analysis of the Empirical Literature. *Pers. Soc. Psychol. Bull.* 21, 1 (January 1995), 58–75. DOI:https://doi.org/10.1177/0146167295211007

[50] Mary Madden and Lee Rainie. 2015. Americans' Attitudes About Privacy, Security and Surveillance | Pew Research Center. Retrieved February 28, 2019 from http://www.pewinternet.org/2015/05/20/americans-attitudes-about-privacy-security-and-surveillance/

[51] Naresh K. Malhotra, Sung S. Kim, and James Agarwal. 2004. Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. *Inf. Syst. Res. Linthicum* 15, 4 (December 2004), 336–355.

[52] Philip Menard, Gregory J. Bott, and Robert E. Crossler. 2017. User Motivations in Protecting Information Security: Protection Motivation Theory Versus Self-Determination Theory. *J. Manag. Inf. Syst.* 34, 4 (October 2017), 1203–1230. DOI:https://doi.org/10.1080/07421222.2017.1394083

[53] Richard G. Netemeyer. 2003. *Scaling procedures : issues and applications*. Sage Publications,.

[54] Kenneth Olmstead and Aaron Smith. 2017. Americans and Cybersecurity. *Pew Research Center: Internet, Science & Tech*. Retrieved November 6, 2017 from http://www.pewinternet.org/2017/01/26/americans-and-cybersecurity/

[55] Will Oremus. 2016. "Is This Something That's Going to Haunt Me the Rest of My Life?" *Slate*. Retrieved September 17, 2018 from http://www.slate.com/articles/technology/future_tense/2016/12/an_interview_with_charles_delavan_the_it_guy_whose_typo_led_to_the_podesta.html

[56] Kathryn Parsons, Dragana Calic, Malcolm Pattinson, Marcus Butavicius, Agata McCormac, and Tara Zwaans. 2017. The Human Aspects of Information Security Questionnaire (HAIS-Q): Two further validation studies.

*Comput. Secur.* 66, (May 2017), 40–51. DOI:https://doi.org/10.1016/j.cose.2017.01.004

[57] Sandy Kristin Piderit. 2000. Rethinking Resistance and Recognizing Ambivalence: A Multidimensional View of Attitudes Toward an Organizational Change. *Acad. Manage. Rev.* 25, 4 (October 2000), 783–794. DOI:https://doi.org/10.5465/amr.2000.3707722

[58] J. O. Prochaska and W. F. Velicer. 1997. The transtheoretical model of health behavior change. *Am. J. Health Promot. AJHP* 12, 1 (October 1997), 38–48.

[59] Elissa M. Redmiles, John P. Dickerson, Krishna P. Gummadi, and Michelle L. Mazurek. 2018. Equitable Security: Optimizing Distribution of Nudges and Resources. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security* (CCS '18), 2270–2272. DOI:https://doi.org/10.1145/3243734.3278507

[60] Elissa M. Redmiles, Sean Kross, and Michelle L. Mazurek. 2016. How I Learned to Be Secure: A Census-Representative Survey of Security Advice Sources and Behavior. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (CCS '16), 666–677. DOI:https://doi.org/10.1145/2976749.2978307

[61] Everett M. Rogers. 2010. *Diffusion of Innovations, 4th Edition*. Simon and Schuster.

[62] Ronald W. Rogers. 1975. A Protection Motivation Theory of Fear Appeals and Attitude Change1. *J. Psychol.* 91, 1 (September 1975), 93–114. DOI:https://doi.org/10.1080/00223980.1975.9915803

[63] Richard M. Ryan and Edward L. Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am. Psychol.* 55, 1 (2000), 68–78. DOI:https://doi.org/10.1037//0003-066X.55.1.68

[64] Yukiko Sawaya, Mahmood Sharif, Nicolas Christin, Ayumu Kubota, Akihiro Nakarai, and Akira Yamada. 2017. Self-Confidence Trumps Knowledge: A Cross-Cultural Study of Security Behavior. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (CHI '17), 2202–2214. DOI:https://doi.org/10.1145/3025453.3025926

[65] Susanne G. Scott and Reginald A. Bruce. 1995. Decision-Making Style: The Development and Assessment of a New Measure. *Educ. Psychol. Meas.* 55, 5 (October 1995), 818–831. DOI:https://doi.org/10.1177/0013164495055005017

[66] Mark Sherer, James E. Maddux, Blaise Mercandante, Steven Prentice-Dunn, Beth Jacobs, and Ronald W. Rogers. 1982. The Self-Efficacy Scale: Construction and Validation. *Psychol. Rep.* 51, 2 (October 1982), 663–671. DOI:https://doi.org/10.2466/pr0.1982.51.2.663

[67] Matthew S. Stanford, Charles W. Mathias, Donald M. Dougherty, Sarah L. Lake, Nathaniel E. Anderson, and Jim H. Patton. 2009. Fifty years of the Barratt Impulsiveness Scale: An update and review. *Personal. Individ. Differ.* 47, 5 (October 2009), 385–395. DOI:https://doi.org/10.1016/j.paid.2009.04.008

[68] Alan Strathman, Faith Gleicher, David S. Boninger, and Scott Edwards. 1994. *The Consideration of Future Consequences: Weighing Immediate and Distant Outcomes of Behavior*. DOI:https://doi.org/10.1037/0022-3514.66.4.742

[69] Stephen Sutton. 1998. Predicting and Explaining Intentions and Behavior: How Well Are We Doing? *J. Appl. Soc. Psychol.* 28, 15 (August 1998), 1317–1338. DOI:https://doi.org/10.1111/j.1559-1816.1998.tb01679.x

[70] M. Tischer, Z. Durumeric, S. Foster, S. Duan, A. Mori, E. Bursztein, and M. Bailey. 2016. Users Really Do Plug in USB Drives They Find. In *2016 IEEE Symposium on Security and Privacy (SP)*, 306–319. DOI:https://doi.org/10.1109/SP.2016.26

[71] Viswanath Venkatesh and Fred D. Davis. 2000. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Manag. Sci.* 46, 2 (2000), 186–204.

[72] Viswanath Venkatesh, Michael G. Morris, Gordon B. Davis, and Fred D. Davis. 2003. User Acceptance of Information Technology: Toward a Unified View. *Manag. Inf. Syst. Q.* 27, 3 (2003), 5.

[73] Rick Wash. 2010. Folk Models of Home Computer Security. In *Proceedings of the Sixth Symposium on Usable Privacy and Security* (SOUPS '10), 11:1–11:16. DOI:https://doi.org/10.1145/1837110.1837125

[74] Rick Wash, Emilee Rader, and Ruthie Berman. 2016. Understanding Password Choices: How Frequently Entered Passwords are Re-used Across Websites. *USENIX Symp. Usable Priv. Secur.* (2016), 15.

[75] Thomas Llewelyn Webb, Paschal Sheeran, Thomas L. Webb, and Paschal Sheeran. 2006. Does changing behavioral intentions engender behavioral change? A meta-analysis of the experimental evidence. In *at PENNSYLVANIA STATE UNIV on September 19, 2016adh.sagepub.comDownloaded from Zigarmi and Nimon 15*, 249–268.

[76] G. L. Zimmerman, C. G. Olsen, and M. F. Bosworth. 2000. A "stages of change" approach to helping patients change behavior. *Am. Fam. Physician* 61, 5 (March 2000), 1409–1416.

[77] 2018. It's Time to Do Something: Mitigating the Negative Impacts of Computing Through a Change to the Peer Review Process. *ACM FCA*. Retrieved September 17, 2018 from https://acm-fca.org/2018/03/29/negativeimpacts/

[78] 2018 Data Breach Investigations Report. *Verizon Enterprise Solutions*. Retrieved April 13, 2018 from http://www.verizonenterprise.com/verizon-insights-lab/dbir/

[79] The Perfect Weapon: How Russian Cyberpower Invaded the U.S. - The New York Times. Retrieved September 17, 2018 from https://www.nytimes.com/2016/12/13/us/politics/russia-hack-election-dnc.html

[80] 2018 Federal Poverty Level Guidelines (FPL): 2018 LIS Qualifications and Benefits. Retrieved February 22, 2019 from https://q1medicare.com/q1group/MedicareAdvantagePartD/Blog.php?blog=2018-Federal-Poverty-Level-Guidelines--FPL---2018-2019-LIS-Qualifications-and-Benefits&blog_id=674&category_id=8

# 12. Appendices

### 12.1. Table of Measures Used in this Report

| *Measures used in this report* | *Rationale for including* |
|---|---|
| SeBIS scale, 16 items [33] | Test correlation with SA-6 |
| Recalled Security Actions, 10 items | Test variances with SA-6 |
| Internet Know-How, 9 items [48] | Test correlation with SA-6 |
| Technical Know-How, 9 items [48] | Test correlation with SA-6 |
| IUIPC scale, 10 items [51] | Test correlation with SA-6 |
| Frequency of falling victim to a security breach, 2 items * | Test variances with SA-6 |
| Amount heard or seen about security breaches, 1 item * | Test variances with SA-6 |
| Whether respondent's security behavior is influenced by other factors or strategies, 1 item | Doubles as attention check; participants directed to leave an answer or type in "None." |
| Barratt Impulsiveness Scale, 30 items [67] | Test correlation with SA-6 |
| Privacy Concern Scale, 16 items [13] | Test correlation with SA-6 |
| Ten-Item Personality Inventory, 10 items [42] | Test correlation with SA-6 |
| General Self-Efficacy scale, 11 items [76] | Test correlation with SA-6 |
| Social Self-Efficacy scale, 5 items [76] | Test correlation with SA-6 |
| Confidence in Using Computers, 12 items [38]** | Test correlation with SA-6 |
| Web-Oriented Digital Literacy, 25 items [44]*** | Test correlation with SA-6 |
| Need for Cognition scale, 18 items [14] | Test correlation with SA-6 |
| GDMS Avoidance and Dependence subscales, 10 items [65] | Test correlation with SA-6 |
| DoSpeRT Health/Safety subscales, 12 items [11] | Test correlation with SA-6 |
| Consideration of Future Consequences scale, 12 items [68] | Test correlation with SA-6 |
| Age range, 1 item **** | Test variances in SA-6 |
| Gender, 1 item **** | Test variances in SA-6 |
| Level of formal education, 1 item **** | Test variances in SA-6 |
| Household income level, 1 item **** | Test variances in SA-6 |
| Employment status, 1 item | Test variances in SA-6 |

*reworded from IUIPC survey **reworded item 12 from original scale ***cut down from 43 items in original scale ****worded to be comparable with Pew surveys

### 12.2. List of Candidate Items for Scale Finalizing

The following is the selected list of $n$=48 candidate items for SA-6 (chosen items are shaded), along with the sources of the items. These were deployed in questionnaires on Amazon Mechanical Turk, the university-run study pool and to the Qualtrics U.S. Census-tailored panel.

| *Candidate items (n=48) analyzed for scale* | *Source* |
|---|---|
| A security breach, if one occurs, is not likely to cause significant harm to my online identity or accounts. | [1,20,21,23] |
| Generally, I am aware of existing security threats. | [20–23] |
| Generally, I am willing to spend money to use security measures that counteract the threats that are relevant to me. | [21,45] |
| Generally, I care about security and privacy threats. | [20–23] |
| Generally, I diligently follow a routine about security practices. | [author generated] |
| Generally, I know how to figure out if an email was sent by a scam artist. | [20] |
| Generally, I know how to use security measures to counteract the threats that are relevant to me. | [20–23] |
| Generally, I know which security threats are relevant to me. | [20–23] |
| Generally, I want to use measures that can counteract security and privacy threats. | [20–23] |
| I always pay attention to experts' advice about the steps I need to take to keep my online data and accounts safe. | [15,21] |
| I always trust experts' recommendations about security measures (such as using unique passwords or a password manager, installing recommended software updates, etc.). | [15,21] |
| I am confident that I am taking the necessary steps to keep my online data and accounts safe. | [20–23] |
| I am confident that I can change my security behaviors, if needed, to protect myself against threats (such as phishing, computer viruses, identity theft, password hacking) that are a danger to my online data and accounts. | [76] |
| I am confident that I could change my security behaviors if I decided to. | [76] |
| I am extremely knowledgeable about all the steps needed to keep my online data and accounts safe. | [20–23] |
| I am extremely knowledgeable about how to take the necessary steps to keep my online data and accounts safe. | [20–23] |
| I am extremely knowledgeable about which security threats (such as phishing, computer viruses, malware, password hacking) are a danger to my online data and accounts. | [20–23] |
| I am extremely motivated to take all the steps needed to keep my online data and accounts safe. | [20–23] |
| I am extremely well aware of existing security threats (such as phishing, computer viruses, identity theft, password hacking). | [20–23] |
| I am extremely well aware of the necessary steps that I can take to counteract security threats (such as phishing, computer viruses, identity theft, password hacking). | [20–23] |
| I am too busy to put in the effort needed to change my security behaviors. | [21,29] |

| | |
|---|---|
| I care very much about the issue of security threats (such as phishing, computer viruses, identity theft, password hacking). | [20–23] |
| I dread that using recommended security measures will backfire on me (such as forgetting a needed password, updated software becoming unusable, etc.). | [21,45] |
| I feel guilty when I do not use recommended security measures (such as by reusing passwords, putting off software updates, etc.). | [21] |
| I generally am aware of existing security measures that I can use to counteract security threats. | [20–23] |
| I generally am aware of methods to send email or text messages that can't be spied on. | [20–23] |
| I have much bigger problems than my risk of a security breach. | [21,29] |
| I need to change my security behaviors to improve my protection against security threats (such as phishing, computer viruses, identity theft, password hacking). | [20,76] |
| I often am interested in articles about security threats. | [24] |
| I seek out opportunities to learn about security measures that are relevant to me. | [21] |
| I usually will not use security measures if they are inconvenient. | [20–23] |
| I usually will not use security measures unless I am forced to. | [20–23] |
| I want to change my security behaviors in order to keep my online data and accounts safe. | [20,76] |
| I want to change my security behaviors to improve my protection against threats (such as phishing, computer viruses, identity theft, password hacking) that are a danger to my online data and accounts. | [20,76] |
| I worry that I'm not doing enough to protect myself against threats (such as phishing, computer viruses, identity theft, password hacking) that are a danger to my online data and accounts. | [20,62] |
| It is a lost cause to take all the steps needed to keep my online data and accounts safe. | [author generated] |
| It is important for me to change my security behaviors to improve my protection against security threats (such as phishing, computer viruses, identity theft, password hacking). | [20,76] |
| It is not possible for me to do more than I already am to counteract security threats (such as phishing, computer viruses, identity theft, password hacking) that are a danger to my online data and accounts. | [author generated] |
| It's a sign of paranoia to use numerous security measures to protect against threats. | [21,41] |
| It's a sign of paranoia to use recommended security measures (such as using unique passwords or a password manager, installing recommended software updates, etc.). | [21,41] |
| My current lapses in using security measures are harmless. | [1,21] |
| My own actions can make a significant difference in keeping my online data and accounts safe. | [10] |
| Oftentimes, as soon as I discover a security problem, I report it to someone who can fix it. | [33] |
| Oftentimes, I am running on "automatic pilot" when I sift through my email and text messages. | [author generated] |
| Oftentimes, I will check that my anti-virus software has been regularly updating itself. | [33] |
| The exposure of my online data and accounts in a security incident, if one occurs, would be a significant problem for me. | [author generated] |
| The theft of my online data or accounts in a security breach, if one occurs, would be a significant problem for me. | [author generated] |
| There are good reasons why I do not take the necessary steps to keep my online data and accounts safe. | [21] |

### 12.3. Pairwise Comparisons for MTurk and University Samples

The following table contains the chi-square statistics for all of the age-level pairwise comparisons (*1=18-29, 2=30-39, 3=40-49, 4=50-59, 5=60 or older; adj. p <.005*). No pairwise comparisons were statistically significant.

| *Pair* | *N* | *df* | *X^2* | *p* |
|---|---|---|---|---|
| **1 vs. 2** | 377 | 1 | 2.88 | (n.s.) |
| **1 vs. 3** | 271 | 1 | 2.39 | (n.s.) |
| **1 vs. 4** | 257 | 1 | 6.44 | (n.s.) |
| **1 vs. 5** | 242 | 1 | 1.26 | (n.s.) |
| **2 vs. 3** | 202 | 1 | 0.18 | (n.s.) |
| **2 vs. 4** | 188 | 1 | 2.48 | (n.s.) |
| **2 vs. 5** | 173 | 1 | 3.41 | (n.s.) |
| **3 vs. 4** | 82 | 1 | 1.10 | (n.s.) |
| **3 vs. 5** | 67 | 1 | 3.62 | (n.s.) |
| **4 vs. 5** | 53 | 1 | 6.86 | (n.s.) |

The following table contains the chi-square statistics for all of the education-level pairwise comparisons (*1=Some high school, 2=High school degree or equivalent, 3=Some college, technical degree or associate's degree, 4=Bachelor's degree, 5=Graduate or professional degree; adj. p <.005*). Some pairwise comparisons were statistically significant, with the sample from the university-run study pool skewing toward higher levels of educational attainment.

| *Pair* | *N* | *df* | *X^2* | *p* |
|---|---|---|---|---|
| **1 vs. 2** | 38 | 1 | 8.16 | **0.004** |
| **1 vs. 3** | 140 | 1 | 0.86 | (n.s.) |
| **1 vs. 4** | 212 | 1 | 0.64 | (n.s.) |
| **1 vs. 5** | 242 | 1 | 1.26 | (n.s.) |
| **2 vs. 3** | 172 | 1 | 12.44 | **0.001** |
| **2 vs. 4** | 144 | 1 | 15.48 | **0.001** |
| **2 vs. 5** | 129 | 1 | 33.6 | **0.001** |
| **3 vs. 4** | 346 | 1 | 0.39 | (n.s.) |
| **3 vs. 5** | 231 | 1 | 14.86 | **0.001** |
| **4 vs. 5** | 303 | 1 | 13.03 | **0.001** |

The following table contains the chi-square statistics for all of the income-level pairwise comparisons (*1=Under $25,000, 2=$25K to $49,999, 3=$50K to $74,999, 4=$75K to $99,999, 5=$100K or higher; adj. p <.005*). Some pairwise comparisons were statistically significant, with the sample from the university-run study pool skewing toward higher levels of yearly household income.

| *Pair* | *N* | *df* | *X^2* | *p* |
|---|---|---|---|---|
| **1 vs. 2** | 256 | 1 | 2.79 | (n.s.) |
| **1 vs. 3** | 207 | 1 | 5.35 | (n.s.) |
| **1 vs. 4** | 180 | 1 | 0.02 | (n.s.) |
| **1 vs. 5** | 186 | 1 | 11.44 | **0.001** |
| **2 vs. 3** | 229 | 1 | 0.75 | (n.s.) |
| **2 vs. 4** | 202 | 1 | 1.58 | (n.s.) |
| **2 vs. 5** | 208 | 1 | 23.62 | **0.001** |
| **3 vs. 4** | 153 | 1 | 3.53 | (n.s.) |
| **3 vs. 5** | 159 | 1 | 26.73 | **0.001** |
| **4 vs. 5** | 132 | 1 | 9.58 | **0.002** |

The following table contains the chi-square statistics for all of the frequency-level pairwise comparisons for personal experiences of a security breach (*1=Very infrequently, 2=Infrequently, 3=Neither infrequently nor frequently, 4=Frequently, 5=Very frequently; adj. p <.005*). No pairwise comparisons were statistically significant.

| *Pair* | *N* | *df* | *X^2* | *p* |
|---|---|---|---|---|
| **1 vs. 2** | 374 | 1 | 0.04 | (n.s.) |
| **1 vs. 3** | 301 | 1 | 2.16 | (n.s.) |
| **1 vs. 4** | 284 | 1 | 1.01 | (n.s.) |
| **1 vs. 5** | 254 | 1 | 0.44 | (n.s.) |
| **2 vs. 3** | 185 | 1 | 2.24 | (n.s.) |
| **2 vs. 4** | 168 | 1 | 0.70 | (n.s.) |
| **2 vs. 5** | 138 | 1 | 0.35 | (n.s.) |
| **3 vs. 4** | 95 | 1 | 3.49 | (n.s.) |
| **3 vs. 5** | 65 | 1 | 1.51 | (n.s.) |
| **4 vs. 5** | 48 | 1 | 0.02 | (n.s.) |

The following table contains the chi-square statistics for all of the frequency-level pairwise comparisons for a close tie's experiences of a security breach (*1=Very infrequently, 2=Infrequently, 3=Neither infrequently nor frequently, 4=Frequently, 5=Very frequently; adj. p <.005*). No pairwise comparisons were statistically significant.

| *Pair* | *N* | *df* | *X^2* | *p* |
|---|---|---|---|---|
| **1 vs. 2** | 329 | 1 | 1.14 | (n.s.) |
| **1 vs. 3** | 223 | 1 | 3.05 | (n.s.) |
| **1 vs. 4** | 191 | 1 | 1.45 | (n.s.) |
| **1 vs. 5** | 140 | 1 | 3.46 | (n.s.) |
| **2 vs. 3** | 282 | 1 | 0.87 | (n.s.) |
| **2 vs. 4** | 250 | 1 | 4.12 | (n.s.) |
| **2 vs. 5** | 199 | 1 | 4.41 | (n.s.) |
| **3 vs. 4** | 144 | 1 | 6.25 | (n.s.) |
| **3 vs. 5** | 93 | 1 | 5.40 | (n.s.) |
| **4 vs. 5** | 61 | 1 | 2.28 | (n.s.) |

The following table contains the chi-square statistics for all of the amount-level pairwise comparisons for what participants have heard or seen about online security breaches (*1=None at all, 2=A little, 3=A moderate amount, 4=A lot, 5=A great deal; adj. p <.005*). No pairwise comparisons were statistically significant.

| *Pair* | *N* | *df* | *X^2* | *p* |
|---|---|---|---|---|
| **1 vs. 2** | 179 | 1 | 1.45 | (n.s.) |
| **1 vs. 3** | 193 | 1 | 1.95 | (n.s.) |
| **1 vs. 4** | 115 | 1 | 1.34 | (n.s.) |
| **1 vs. 5** | 63 | 1 | 2.36 | (n.s.) |
| **2 vs. 3** | 324 | 1 | 0.13 | (n.s.) |
| **2 vs. 4** | 246 | 1 | 0.00 | (n.s.) |
| **2 vs. 5** | 194 | 1 | 0.57 | (n.s.) |
| **3 vs. 4** | 260 | 1 | 0.08 | (n.s.) |
| **3 vs. 5** | 208 | 1 | 0.29 | (n.s.) |
| **4 vs. 5** | 130 | 1 | 0.48 | (n.s.) |

### 12.4. Factor Loadings, Alpha if Item Deleted and Item Histograms for SA-6 Scale Finalization (n=478)

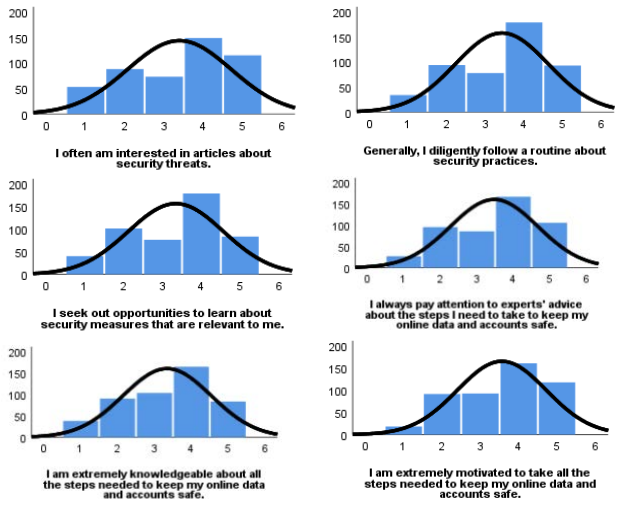
***Principal Component Analysis - Factor Loading***

| | |
|---|---|
| I always pay attention to experts' advice about the steps I need to take to keep my online data and accounts safe. | 0.84 |
| Generally, I diligently follow a routine about security practices. | 0.82 |
| I seek out opportunities to learn about security measures that are relevant to me. | 0.81 |
| I am extremely motivated to take all the steps needed to keep my online data and accounts safe. | 0.80 |
| I often am interested in articles about security threats. | 0.79 |
| I am extremely knowledgeable about all the steps needed to keep my online data and accounts safe. | 0.74 |

***Cronbach's Alpha if Item Deleted (overall = .89)***

| | |
|---|---|
| I often am interested in articles about security threats. | 0.87 |
| Generally, I diligently follow a routine about security practices. | 0.86 |
| I seek out opportunities to learn about security measures that are relevant to me. | 0.86 |
| I always pay attention to experts' advice about the steps I need to take to keep my online data and accounts safe. | 0.86 |
| I am extremely knowledgeable about all the steps needed to keep my online data and accounts safe. | 0.88 |
| I am extremely motivated to take all the steps needed to keep my online data and accounts safe. | 0.87 |

***Histograms (1=Strongly disagree, 5=Strongly agree)***




### 12.5. Cronbach's alpha for composite measures incl. for convergent and discriminant validity (threshold = .70)

| *Measure* | *Alpha* |
|---|---|
| Barratt Impulsivity Scale | **0.86** |
| Big5-Agreeableness | 0.40 |
| Big5-Conscientiousness | 0.53 |
| Big5-Emotional Stability | 0.60 |
| Big5-Extraversion | **0.70** |
| Big5-Openness to Experiences | 0.37 |
| Confidence in Using Computers | **0.89** |
| Consideration of Future Consequences | **0.77** |
| DoSpERT - Risk-perception subscale | **0.89** |
| DoSpERT - Risk-taking subscale | **0.84** |
| GDMS – Avoidance subscale | **0.91** |
| GDMS – Dependence subscale | **0.81** |
| Kang Internet Know-How scale | **0.91** |
| Kang Technical Know-How scale | 0.63 |
| Need for Cognition scale | **0.88** |
| Privacy – Internet Users' Infor. Privacy Concerns | **0.88** |
| Privacy – Privacy Concerns Scale | **0.96** |
| SeBIS – Security Behavior Intentions Scale | **0.70** |
| Self-Efficacy - General | **0.90** |
| Self-Efficacy - Social | **0.75** |
| Web-oriented Digital Literacy | **0.94** |

**12.6. Classification Tables for Selected Logistic Regressions**

Q24_4b "*In the past week, I have downloaded and installed at least one available update for my computer's operating system within 24 hours of receiving a notification that it was available*":

Predicted – Constant only

| Observed | | Q24_4b 1.00 | Q24_4b 2.00 | Percentage Correct |
|---|---|---|---|---|
| Q24_4b | 1.00 | 0 | 79 | .0 |
| | 2.00 | 0 | 112 | 100.0 |
| Overall Percentage | | | | 58.6 |

Predicted – SeBIS only

| Observed | | Q24_4b 1.00 | Q24_4b 2.00 | Percentage Correct |
|---|---|---|---|---|
| Q24_4b | 1.00 | 41 | 38 | 51.9 |
| | 2.00 | 24 | 88 | 78.6 |
| Overall Percentage | | | | 67.5 |

Predicted – SeBIS with SA-6

| Observed | | Q24_4b 1.00 | Q24_4b 2.00 | Percentage Correct |
|---|---|---|---|---|
| Q24_4b | 1.00 | 45 | 34 | 57.0 |
| | 2.00 | 27 | 85 | 75.9 |
| Overall Percentage | | | | 68.1 |

Q24_7b "*In the past week, I have verified at least once that I am running antivirus software that is fully updated*":

Predicted – Constant only

| Observed | | Q24_7b 1.00 | Q24_7b 2.00 | Percentage Correct |
|---|---|---|---|---|
| Q24_7b | 1.00 | 108 | 0 | 100.0 |
| | 2.00 | 97 | 0 | .0 |
| Overall Percentage | | | | 52.7 |

Predicted – SeBIS only

| Observed | | Q24_7b 1.00 | Q24_7b 2.00 | Percentage Correct |
|---|---|---|---|---|
| Q24_7b | 1.00 | 74 | 34 | 68.5 |
| | 2.00 | 38 | 59 | 60.8 |
| Overall Percentage | | | | 64.9 |

Predicted – SeBIS with SA-6

| Observed | | Q24_7b 1.00 | Q24_7b 2.00 | Percentage Correct |
|---|---|---|---|---|
| Q24_7b | 1.00 | 72 | 36 | 66.7 |
| | 2.00 | 34 | 63 | 64.9 |
| Overall Percentage | | | | 65.9 |

Q24_10b *"In the past week, I have used a password/passcode at least once to unlock my tablet"*:

Predicted – Constant only

| Observed | | Q24_10b 1.00 | Q24_10b 2.00 | Percentage Correct |
|---|---|---|---|---|
| Q24_10b | 1.00 | 136 | 0 | 100.0 |
| | 2.00 | 57 | 0 | .0 |
| Overall Percentage | | | | 70.5 |

Predicted – SeBIS only

| Observed | | Q24_10b 1.00 | Q24_10b 2.00 | Percentage Correct |
|---|---|---|---|---|
| Q24_10b | 1.00 | 125 | 11 | 91.9 |
| | 2.00 | 34 | 23 | 40.4 |
| Overall Percentage | | | | 76.7 |

Predicted – SeBIS with SA-6

| Observed | | Q24_10b 1.00 | Q24_10b 2.00 | Percentage Correct |
|---|---|---|---|---|
| Q24_10b | 1.00 | 125 | 11 | 91.9 |
| | 2.00 | 33 | 24 | 42.1 |
| Overall Percentage | | | | 77.2 |

# The Effect of Entertainment Media on Mental Models of Computer Security

Kelsey R. Fulton, Rebecca Gelles, Alexandra McKay,
Richard Roberts, Yasmin Abdi, and Michelle L. Mazurek
*University of Maryland*
*{kfulton, rgelles, ricro, mmazurek}@cs.umd.edu*
*{amckay12, yabdi}@terpmail.umd.edu*

## Abstract

When people inevitably need to make decisions about their computer-security posture, they rely on their mental models of threats and potential targets. Research has demonstrated that these mental models, which are often incomplete or incorrect, are informed in part by fictional portrayals in television and film. Inspired by prior research in public health demonstrating that efforts to ensure accuracy in the portrayal of medical situations has had an overall positive effect on public medical knowledge, we explore the relationship between computer security and fictional television and film. We report on a semi-structured interview study (n=19) investigating what users have learned about computer security from mass media and how they evaluate what is and is not realistic within fictional portrayals. In addition to confirming prior findings that television and film shape users' mental models of security, we identify specific misconceptions that appear to align directly with common fictional tropes. We identify specific proxies that people use to evaluate realism and examine how they influence these misconceptions. We conclude with recommendations for security researchers as well as creators of fictional media when considering how to improve people's understanding of computer-security concepts and behaviors.

## 1 Introduction

Computer users frequently make security-relevant decisions during password creation, link navigation, messaging platform selection, and other activities. These choices reflect users' mental models about what is risky, what is safe, and how computer systems work. However, many people have limited knowledge of computer security or computers generally; users' mental models are often incomplete or incorrect [2, 27]. Erroneous mental models can lead users to inaccurate conclusions about how to best protect themselves online (e.g., believing that standard text messages are safer than encrypted chat messages) [2].

Prior studies have shown that mass media, including television and film, can influence user mental models of computer security [21–23]. This phenomenon has been observed in other fields as well: The portrayal of medical information in television and film, and its effect on viewers, has been studied extensively, leading to concrete efforts to improve the accuracy of medical information shown to the public. Programs like "Hollywood, Health and Society" provide consultation to the entertainment industry to help ensure that fictional medical storylines are accurate and avoid disseminating harmful disinformation [1]. Research suggests that, overall, mass-media portrayals have imperfect but positive effects on viewers' medical knowledge [12].

In contrast, we are aware of no similar effort to improve accuracy in mass-media depictions of computer security. Depictions of computer security and "hacking" in mass media vary, but are often unrealistic, including confusing jargon, unnecessary visuals of internal computer operations, rapid hacking and counter-hacking, and other tropes [7, 16]. There has been little or no research effort to understand how these portrayals affect users' security beliefs and behaviors. Merely exposing users to the concept of computer security may improve their understanding or awareness. However, inaccurate and exaggerated portrayals could also harm development of healthy mental models.

To investigate this question, we conducted a semi-structured interview study (n=19) to gauge how media portrayals affect people's perceptions of computer security and hackers as well as their resulting mental models. We asked participants broadly about their prior computer security knowledge, experience, and mass media background. We then showed each participant six clips involving computer security from


*USENIX Symposium on Usable Privacy and Security (SOUPS)* 2019.
August 11–13, 2019, Santa Clara, CA, USA.

television shows and movies, chosen to depict different technological and social dimensions both more and less accurately, asking participants to evaluate the realism of each clip and to explain their reasoning and judgment. In particular, we focused on three main research questions:

RQ1. What do people learn about computer security from mass-media portrayals?

RQ2. How do these learned concepts affect people's overall mental models of computer security and their resulting security behaviors?

RQ3. Why do people learn these particular concepts: for example, why do they find certain portrayals more believable or compelling?

We found that mental models were often affected by incorrect or incomplete information presented in fictional media. While incorrect low-level technical details may not be inherently detrimental, misunderstandings of high-level technical takeaways could cause harm. Media portrayals teach and/or reinforce several common mental models observed in previous work, some of which play a factor in non-optimal security decisions made by users. These include the beliefs that security intrusions are always obvious, that precautions are pointless because hacking is inevitable, and that ordinary users are not important enough to be hacked. Not everything learned from fictional media was negative: participants gained awareness of the danger of phishing and suspicious emails.

We also discovered that participants' trust in fictional depictions — and willingness to incorporate them into their mental models — depends on several key factors. Most participants begin with a default judgment about the average realism of media portrayals (on any topic) and adjust this perception based on cues including their own technical knowledge, their ability to relate depicted events to their own experiences, and the cinematic qualities of the scene. Our results suggest a feedback loop; perceived realism is tied to conformance with pre-existing mental models of security, which may arise in part from prior exposure to fictional media.

We suggest several avenues for improvement. Entertainers should take more responsibility to mitigate the spread of misinformation. Researchers and educators can use our results to better understand the assumptions and decisions that users make when performing security-relevant behaviors and incorporate that knowledge into the design of tools, interventions, and educational content. Perhaps there is room for collaboration toward a common goal of educating users while still keeping them engaged and entertained.

## 2 Background & related work

Prior work has examined the sources of people's computer-security knowledge. Rader et al. discussed how stories convey security knowledge between individuals, finding this can be an effective medium to change user thought and behavior [20]. In this work, we explore the question of whether stories told in fictional media, rather than person-to-person interaction, also affect knowledge and behavior.

Redmiles et al. investigated why users accept or reject security advice [22], concluding that negative events often motivate users to take advice from fictional and nonfictional stories. In follow-up work, Redmiles et al. found that 67.5% of a broad U.S. sample cited media as a source of security knowledge, and 25% of those participants specifically referenced fictional narratives [21]. Ruoti et al. also found that participants learned about security threats from media [23], and Forget et al. established that many users outsource their decisions to trusted experts, including media [5]. Our work expands on these findings by examining what people learn specifically from fictional television and movies; we explore how perceived realism in fictional media shapes security beliefs.

Other researchers have explored users' mental models of security in detail. Kang et al. used a drawing exercise to elicit mental models of the internet as a whole, finding surprisingly similar security beliefs among users with different levels of technical knowledge [13]. These included beliefs that average users were not of serious interest to hackers, and that security efforts are inherently futile because attackers are too powerful. These beliefs align with those identified by Wash and Rader, who classified several common folk models of attackers as focusing on high-value targets like government officials and wealthy people [27, 28]. Similar mental models were also identified by Abu-Salma et al. in the context of secure messaging applications [2]. This research also identified several misconceptions about the relative security of different communications mechanisms, and about encryption more broadly. Mistaken mental models of encryption were also identified by Wu and Zappala [30]. This broad confirmation of mistaken mental models underscores the importance of understanding where these misconceptions come from.

Prior research has also investigated how mistaken mental models can influence security decision making. Acquisti et al. explored the privacy paradox, in which users express concerns about privacy but do not act to mitigate privacy threats [3]. The authors postulate that this disconnect is due in part to misinformation, as many participants could not correctly identify the likelihood of various privacy abuses. This paper, as well as another by Herley, argues that users trade off the cost of compliance with recommended security behaviors with their perception of the associated reduction in risk [11]. Herley notes that users become habituated to dangers that are oversold or exaggerated. If media suggests that protective behaviors are unnecessarily complicated or onerous, or that threats are so powerful that they cannot be averted, users' willingness to comply may be reduced. Similarly, when media presents exaggerated threats, users' ability to recognize less

dramatic security problems, such as those encountered in real life, may be impaired. We explore these hypotheses further in this work.

The effect of fictional media on behavior has been explored in other contexts, such as portrayals of medical conditions. Whittier et al. found that those who viewed a storyline about syphilis in a fictional TV show were more likely to report intention to be screened, and to inform others about the risk [29]. The effect is so pronounced that in 1997 the Centers for Disease Control and Prevention launched the "Health, Hollywood, and Society" program, to provide medical information to writers in Hollywood [1]. It is clear that portrayals in fictional media can and do affect people's behavior, both positively and negatively. We apply a similar frame to portrayals of computer security.

## 3 Method

To understand how fictional TV shows and movies inform viewers' mental models, we conducted semi-structured interviews with people in the Washington, DC area. In this section we detail our recruitment, the content of our interview protocol (including piloting), our clip selection, our data analysis approach, ethical considerations, and limitations of our study.

### 3.1 Recruitment

To recruit participants, we posted an advertisement to Craigslist in the DC region offering $30 as compensation for participation in an hour-long interview study. Those who wanted to participate were directed to a pre-screening survey. After consenting to the survey, participants entered general demographic information, provided an email address for future contacting, and answered questions about the amount and type of media they consume. From those responses, we selected a sample to invite for interviews, using a modified first-come, first-served approach that also considered demographics. In particular, we selected for diversity of ages, educational backgrounds, and ethnicities, as well as in self-reported frequency of media watching and preferred genres. We also focused on selecting participants with limited technical backgrounds.

We interviewed participants until we stopped hearing substantially new ideas, resulting in a total of 19 participants. This approach was validated when no new codes were created while analyzing any of the final three participants. This sample size aligns with qualitative best practices [8].

### 3.2 Interview protocol

During September and October 2018, we conducted 19 in-person, semi-structured interviews on the University of Maryland campus. Each session lasted about an hour. Most interviews were conducted by two interviewers; due to scheduling conflicts two interviews were conducted solo. Each interview was audio recorded, with permission.

| ID | Gender | Age | Ethn. | Educ. | TV hrs |
|---|---|---|---|---|---|
| P1 | M | 30-39 | B | HS | 50 |
| P2 | F | 30-39 | B | AD | 8 |
| P3 | M | 30-39 | AHP | BD | 14 |
| P4 | F | 40-49 | W | BD | 15 |
| P5 | F | 30-39 | B | HS | 15 |
| P6 | F | 50-59 | B | SC | 35 |
| P7 | F | 18-29 | W | BD | 10 |
| P8 | M | 40-49 | HL | BD | 20 |
| P9 | F | 50-59 | O | PD | 2 |
| P10 | M | 30-39 | B | SC | 3 |
| P11 | M | 60-69 | B | SC | 20 |
| P12 | F | 30-39 | B | MD | 5 |
| P13 | F | 18-29 | W | HS | 8 |
| P14 | F | 18-29 | B | BD | 6 |
| P15 | F | 18-29 | B | BD | 3 |
| P16 | M | 18-29 | B | SC | 4 |
| P17 | F | 60-69 | W | BD | 15 |
| P18 | M | 50-59 | B | HS | 6 |
| P19 | M | 18-29 | HL | SC | 20 |

**Gender:** F - Female, M - Male
**Ethnicity:** AHP - Asian/Native Hawaiian/Pacific Islander, B - Black/African American, HL - Hispanic/Latino, W - White/Caucasian, O - Other
**Education:** HS - High school graduate/diploma/equivalent, SC - Some college credit (no degree), AD - Associate's Degree BD - Bachelor's degree, MD - Master's Degree, PD - Professional Degree

Table 1: Participant demographic information including gender, age, ethnicity, educational attainment, and estimated average hours of TV watched per week.

The interview protocol had three phases. Phase one assessed interviewees' familiarity with computer security topics, exposure to security breaches, and pre-existing notions of the portrayal of security in media. First, they participated in a word association exercise. Participants were given the words "cybersecurity," "hacker," and "encryption" and encouraged to define the words or respond with any other related terms that came to mind. Next, participants were asked questions about hackers' goals, capabilities, and limitations. They were then encouraged to talk about a time they or someone they knew had been a victim of hacking, and finally, they were asked to describe times they had seen a depiction of hacking or cyber security portrayed in fictional media. The goal of this phase was to understand participants' mental models before showing them media clips that might influence their answers.

In phase two, participants were shown six scenes from television and movies depicting computer security topics. After each clip, participants were asked if they had any prior exposure to the clip or its source. Next, to assess comprehension, they were asked to summarize the scene. Finally, interviewees were asked to describe which parts of the scene were realistic and unrealistic and why they felt this way. Participants who gave overly general responses were prompted to assess the realism of specific aspects of each scene. The order of the

clips was randomized for each participant to mitigate ordering bias. We describe the six clips we used in Section 3.3.

The final phase dealt with the relative realism of security portrayals in fictional television and movies as a whole. Participants were asked how accurately TV and movies in general portray cybersecurity, hackers, and cryptography. They were then asked if there were any shows or movies that they felt portrayed those topics particularly realistically or unrealistically. Finally, interviewees were encouraged to share any additional thoughts on the subjects discussed in the interview. The full interview protocol is given in Appendix B.

Prior to the 19 main interviews, we conducted a formal pilot with five participants to test our initial interview script and selected media clips. Based on the results, we adjusted our choice of clips since some proved to be difficult to understand out of context. Because the pilot interviews were shorter than initially anticipated, we added a sixth clip (having piloted with five) to increase variety. We also improved some unclear question wording and increased the scope of two questions.

## 3.3 Clips

We selected six video clips from TV shows and movies, sourced from the research team's background knowledge, discussions with peers and colleagues, and online collections of computer science in media [26]. We carefully selected these clips to cover a broad spectrum of hacking scenarios, tropes, realism, and alignment with "folk models" of hacking identified in prior work [27]; these selection criteria are outlined in Table 2. We summarize the six clips below.

**Superman 3.** (1983): A man (Richard Pryor) receives his first paycheck at work and is disappointed to learn how much is taken out in taxes. A co-worker points out he's probably making a half cent more, and that in large corporations there are often fractions left over. When pressed, he admits he doesn't know where that money goes, but the computers probably do. Pryor's character then "hacks" into the system to re-route all the half cents into his account. To do this, he types English sentences, like "Override all security," into a black and green terminal. We chose this clip for its portrayal of a financial motivation and for the unrealistic nature of the hacking.

**NCIS S2E4, "The Bone Yard."** (2004): Members of the NCIS team realize their computer is being hacked when its screen rapidly flashes various windows, images, and code snippets. Two members type on the same keyboard simultaneously to fend off the attacker, but do not succeed as the hacker breaks through "DoD Level 9 Encryption." The computer screen goes dark, and it is revealed that a third team member unplugged the computer to end the attack. This clip was an inspiration for this research project. We chose this clip for its depiction of rapid, real-time hacking and counter-hacking as well as the simple solution to the problem.

| Attribute | Superman 3 | NCIS | Blackhat | Sneakers | Skyfall | Gumball |
|---|---|---|---|---|---|---|
| *Technical Qualities* | | | | | | |
| Realistic Jargon | ○ | ○ | ○ | ● | ○ | ● |
| Unrealistic Jargon | ○ | ● | ○ | ○ | ○ | ○ |
| Realistic Hacker Capabilities | ○ | ○ | ● | ○ | ○ | ● |
| Unrealistic Hacker Capabilities | ○ | ○ | ○ | ● | ○ | ○ |
| Unplugging as a Defense | ○ | ● | ○ | ○ | ◑ | ○ |
| Simultaneous Hacking/Defending | ○ | ● | ○ | ○ | ● | ○ |
| "Flashy"Hacking Visuals | ○ | ● | ○ | ● | ● | ○ |
| Hacking is Obvious to Target | ○ | ● | ○ | ○ | ● | ○ |
| *Type of Hacking* | | | | | | |
| Phishing | ○ | ○ | ● | ○ | ○ | ○ |
| Breaking Cryptography | ○ | ○ | ○ | ● | ○ | ○ |
| Network Intrusion | ○ | ● | ○ | ● | ● | ● |
| Privilege Escalation | ● | ○ | ○ | ● | ○ | ○ |
| *Nontechnical Qualities* | | | | | | |
| Played for Drama | ○ | ○ | ● | ● | ● | ○ |
| Played for Humor | ● | ● | ○ | ○ | ○ | ● |
| Pre-Internet Setting | ● | ○ | ○ | ● | ○ | ○ |
| Post-Internet Setting | ○ | ● | ● | ○ | ● | ● |
| Professional Setting | ● | ● | ● | ○ | ● | ○ |
| Cartoon Animation | ○ | ○ | ○ | ○ | ○ | ● |
| *Who is the Target?* | | | | | | |
| Individual | ○ | ◑ | ● | ○ | ○ | ○ |
| Organization | ● | ◑ | ○ | ● | ● | ● |
| *Who is Hacking?* | | | | | | |
| Protagonist | ● | ● | ● | ● | ● | ● |
| Antagonist | ○ | ● | ○ | ○ | ● | ○ |
| *Folk Models Depicted [27]* | | | | | | |
| Digital Graffiti Artist | ○ | ○ | ○ | ● | ● | ● |
| Burglar | ● | ● | ● | ○ | ○ | ○ |
| Target "Big Fish" | ● | ○ | ○ | ● | ● | ○ |
| Contractor | ○ | ○ | ● | ● | ○ | ○ |
| *Hacker's Goal* | | | | | | |
| Disrupt | ○ | ● | ○ | ○ | ● | ○ |
| Gain Access | ○ | ○ | ○ | ● | ○ | ● |
| Steal | ● | ○ | ● | ○ | ○ | ○ |

○- not present, ◑- partially present/ambiguous, ●- present

Table 2: Evaluation of each clip based selection criteria.

**Blackhat.** (2015): One NSA employee sitting alone at a desk receives an email suggesting he change his password due to his contact with a Joint Task Force. As he downloads a PDF titled "Password Security Guidelines," the scene cuts to two people sitting elsewhere. A man explains to the woman next to him that the PDF the employee has downloaded is actually a keylogger. The scene then cuts between the keyboard on which the employee is typing a new password and the hacker pair's screen, which shows the new password being updated in real time. Both the original and new passwords the NSA employee types are sequences of random letters, special characters, and numbers. The hacker pair successfully log in using the employee's stolen credentials. This scene was chosen for the brevity of the scene and the realism of the hack.

**Sneakers.** (1992): One group of people watches as a man asks another group to list things that are impossible to access, such as major government agencies. He then demonstrates the ability to decrypt information from these agencies via a "chip" that is the "key to unlock everything." As the man accesses the Federal Reserve, "encrypted" gibberish appears on the screen, which the man "decrypts" into useful and meaningful content. He replicates the feat, within a minute, with the national power grid and air traffic control, to the group's astonishment. He explains he has solved the "impossible" mathematical problems that are the root of encryption and hardwired that solution onto a chip. This technical description is in a sense quite accurate, essentially describing a scenario in which an attacker succeeds in, e.g., factoring large numbers in order to break RSA. (Of course, current real-world attackers generally rely instead on much more commonly available software flaws and human errors.) We chose this clip for its portrayal of encryption and the depiction of an attack that was plausibly realistic, but whose realism came from technical knowledge at a depth such that most casual viewers were unlikely to possess it.

**Skyfall.** (2012): A field agent and a technical expert attempt to break into a laptop. The laptop is described as a "polymorphic engine" that changes as though "it's fighting back." The field agent notices a keyword among values being flashed on the screen that, when entered, reveals a map of the London Underground. Doors suddenly fly open, and the technician realizes their computer system has been breached. The breach was a result of plugging a malicious laptop directly into their computing infrastructure, which was the reason this scene was selected. It also features very high-quality graphics and a significant amount of plausible-sounding but incorrect technical language.

**The Amazing World of Gumball S3E32, "The Safety."** (2015): In this cartoon, two characters walk sneakily through an industrial building. When they reach a locked door, the blue creature laments while the pink one says "H-A-C-K hack, press enter" while poking the keyboard. When the door opens, the blue one is shocked, until the pink one launches into an extremely detailed, fast, and fairly realistic diatribe explaining how she did it, which includes references to the VNX array head, decrypting SAS disks, rerouting traffic, and accessing the ESXI server cluster. The scene was selected for its realistic but complex jargon, juxtaposed with the childlike nature of the show and graphics.

## 3.4 Data analysis

Once interviews were complete, we transcribed the audio recordings as a team. After transcription, two team members independently analyzed the data using iterative open coding, developing the codebook incrementally and resolving disagreements after every three transcripts [24]. The two researchers achieved an overall reliability of Krippendorff's $\alpha = 0.75$ [14], calculated using ReCal2 [6]. This level of agreement is above the commonly recommended thresholds of 0.667 [10] or 0.70 [15]. Finally, the research team worked together in an iterative axial coding process to derive larger themes and theories from the fine-grained open codes.

## 3.5 Ethics

Both the initial pilot and the larger main study were approved by University of Maryland's Institutional Review Board. We obtained informed consent before the pre-screening survey and again before the interview. One clip we showed contained strong language. We warned each participant about this and reminded them that they could stop the clip or the entire interview if they felt uncomfortable; none did.

## 3.6 Limitations

As with most qualitative studies, the generalizability of our results is limited by our small sample size. We attempted to mitigate this by recruiting a relatively diverse cohort of participants within the Washington, D.C. area.

We limited our study to television and movies, leaving out other fictional media such as books or podcasts. Similarly, we chose only six clips in order to keep the study short and reduce cognitive burden on interviewees. We sought to choose clips illustrating a variety of ways that cybersecurity is represented in fictional media, but it was not possible to include everything. Nonetheless, we feel that our results provide useful insight across a range of cybersecurity portrayals.

In common with all semi-structured interviews, there is the potential for demand effects, social desirability bias, and satisficing to affect participants' responses. Demand effects, in which participants attempt to provide the answers they believe the interviewer wants to hear, are a particular risk when asking participants to evaluate realism in an area where they may have limited personal knowledge or intuition [18]. To mitigate this, we avoided mentioning the goal of the study directly, and we emphasized to participants that there were no right or wrong answers to our questions.

# 4 Misconceptions derived from TV and film

Participants explicitly connected their beliefs about computer security to fictional portrayals. Five participants did so even without prompting; for example, when asked about the word "hacker," P2 mentioned "Matthew Broderick in WarGames." When asked where they thought their ideas about computer security originated, three participants named TV and movies; others mentioned media more broadly.

P8, who believes media portrayals are generally accurate, said, "The fact at least two movies have [attempts to steal

| Triggering Event | What Users Learn | Influence On User Mindset |
|---|---|---|
| Unplugging the computer stops the hacker | It's easy to recover from being hacked | Failure to follow necessary steps to mitigate damage after an attack |
| Hackers leave call signs or don't hide behavior | If I get hacked, I'll be able to tell | No response to subtle compromises and false assumption of security after missing non-obvious indicators |
| Hackers can break all encryption simultaneously | Protections are weak and security is futile | Lack of implementation of common-sense security practices out of belief they won't make any difference |
| Hackers target particular high-value entities | I'm not important so I won't get hacked | Failure to take precautions out of belief they won't be targets; precautions taken against targeted rather than broad attacks |
| Phishing successfully compromises user accounts* | Suspicious email can lead to being hacked | Greater care taken when evaluating email links and attachments and decreased assumption of security |

Table 3: Examples of Mental Models Drawn from Mass Media (*denotes correct mental model)

missing half-cents from paychecks] probably means someone tried to do it, or did do it." But even those who claimed to believe fictional media is inaccurate, such as P14, drew similar conclusions: when asked what encryption is used for, she struggled for an explanation before mentioning "I think about the movie with . . . " and trailing off. Similarly, P7 said media portrays computer security "probably inaccurately" but also said "sometimes when I'm watching movies or TV and someone is doing like equations. . . I'm like 'Is that real or is that made up?' And that makes me think of a lot of the stuff that I don't really question and I'm just like 'Okay that's what it looks like.' Or I just accept those representations." It appears fictional portrayals may be influencing mental models even for people who claim to know better.

These results align well with findings from Redmiles et al. and Ruoti et al. that fictional media can be a major source of security information for users [21, 23]. Further, our results suggest a kind of feedback loop: people learn mental models of security — sometimes from fictional media — and then these models are reaffirmed when they appear in other media later. We highlight below a few ways in which TV and film portrayals seem to have contributed to our participants' security beliefs. These findings are summarized in Table 3.

## 4.1 Hackers have specific, important targets

When asked about hackers' normal targets, many participants seemed to think hackers only choose important targets, and that targets are always a specific person or entity (rather than, for example, sending a phishing email broadly to many recipients). For example, P18 believes hackers target "very important information, data. The military, law enforcement personnel. . . banks or important corporations, military, intelligence, so they can use it to their advantage." This model was affirmed while watching the Sneakers clip: P18 assumed an attack on "stuff that's been encrypted" would target "intelligence of maybe the Navy or the United States or whatever, banking, stuff like that, it's really huge."

Similarly, P6 commented that she was not important enough to be a target of a hacker: "I'm not a rich person. . . so you know I guess they probably just left me alone." She later connected this belief to the scene in Sneakers, noting that something "that I found realistic was the things that they were breaking the codes to were very high security."

Participants also expected hackers to focus on individual targets rather than random victims. P5, for example, observed that a hacker's target might be "someone that you have a personal grudge against." This mental model was later affirmed when this participant noted that in Blackhat, the targeting of a specific victim personally was "pretty realistic." Participants who believe attacks are targeted specifically rather than at random may choose non-optimal protective behaviors; for example, as Ur et al. demonstrated, users who expect targeted attacks choose passwords differently than than those seek to protect from more general guessing [25].

In general, if people think that only important entities will be attacked and hackers only attack individuals rather than make widespread attacks, they may never worry about protection because they do not view themselves as a potential target. This aligns with Wash's folk model that "hackers are criminals who target big fish" [27].

## 4.2 Attacks and unsafe situations are obvious

People's ability to correctly recognize evidence of security breaches (or conversely, an unfounded fear that they may be under attack) depends on their idea of what security incidents look like. We found that mental models about what hacking looks like were strongly influenced by portrayals in fictional media, in which hacking is commonly portrayed as a dramatic, active intrusion that triggers anomalous behavior, sometimes including a deliberate "signature" from an attacker. Among our clips, this was most noticeable in NCIS, in which an attack led to obvious pop-up windows and messages, and Skyfall, in which defenders were alerted to an intrusion in real time by the attacker's red-skull call sign.

Participants consistently indicated these attack models were realistic. With Skyfall, some participants said that an attacker might not want to tip off the defenders by displaying a call sign; however, P9 specifically mentioned the skull as a good

indicator that the system was under attack. When describing what she found realistic about the NCIS clip, P14 noted, "Especially in the past getting viruses . . . I can remember that happening with a whole bunch of pop-ups in the screen." The clip affirmed this participant's idea that security problems have obvious indicators. P7 summarized these thoughts well: "I feel like imagery of like being hacked where like all the screens flash and stuff is like what is shown in pop culture a lot of times. . . But in my mind that is like, [an indicator that] hacking is happening." If people are waiting for this kind of obvious indicator to identify a security problem, they are likely never to find it.

### 4.3 Encryption is fragile and all security measures are futile

Fictional media commonly portrays encryption as quickly and easily broken by sufficiently talented attackers. This is exemplified by our clip from Sneakers, in which a "master chip" can decrypt data from any secure facility quickly and easily. Most participants found this highly plausible. For example, P19 said, "I feel like we are at the point where people could logistically have that much control over air traffic control, federal reserve, etc. and cause a lot of harm, so that doesn't seem like a far stretch." The clip seems to confirm this participant's existing mental model that security measures may be futile: asked about hackers' limitations, he said "There's always new things to be discovered and with the passage of time technology is only going to get better, so I don't see any limitations whatsoever." P23 echoed this idea, saying that hackers "have no limitations." He connected this idea to the Superman 3 clip: "in our day and age, it's like, in a blink of an eye it's like done, you're not protected."

Further, P6 said that after watching the Sneakers clip, "now I understand what encrypted means.. . . And do I think [the master key's] real? Yes. I do think that is realistic. I think that there is a code among all codes to, I mean maybe not work for everything, but work for a majority of things."

These explanations reflect a fundamental lack of faith in encryption with the potential to engender distrust in products or services that advertise security and privacy. While it may indeed be nearly impossible to stop a sufficiently talented, motivated, and resourced attacker, this mentality could prevent people from taking precautions that could stop many other classes of attackers. These media portrayals are one potential explanation for the finding by Abu-Salma et al. that some users believe secure messaging is not worthwhile because encryption can always be broken by technically savvy attackers who understand it [2].

### 4.4 Unplugging and other solutions

Fictional media often presents simple, facile solutions to complex security problems. Several participants found the NCIS solution — unplugging the computer to end an attack — highly plausible. P16, for example, noted, "The other colleagues came in to help and what solved the situation was pulling the plug. . . . Pull the plug, and it stopped pretty much everything." This affirmed his mental model for solving his own security issues: "There was a time when [hacking] happened, I just pressed the off button really fast to stop it.. . . So the next time I just turned the game off completely. When I turned the whole system off completely, it didn't go down any further. I just turned the whole thing off to stop it."

This sentiment is echoed by P15 who said "But what actually seemed real was when like the dude unplugged it all — cause you know back when I had viruses the first thing I'd do is unplug it and see if it worked again." This kind of straightforward adoption of simple precautions and solutions offered in fictional media, many of which may not be correct, can harm people's efforts to protect themselves.

Also, not one participant mentioned the use of a single standard keyboard by two people simultaneously in the NCIS clip as an unrealistic feature of the scene. While this is a minor detail, not ultimately related to security beliefs and behaviors, it does underscore that viewers take scenes like this one at face value in ways that may be harmful.

### 4.5 Suspicious emails can be dangerous

Not everything portrayed in fictional media is detrimental to peoples' mental models. Many participants noted that receiving a suspicious email is a proxy for getting hacked. P4 mentioned, "You always hear about viruses that can attach something so that's why you never open attachments unless you know who sent it to you" as a reason why the Blackhat clip seemed realistic. Describing the same clip, P11 said, "I don't know very much about phishing except from just what I've heard or read. But I think that's pretty much the way it works.. . . snatch or steal the information. I'm guessing that that looks legitimate; that looks real." In these cases, the media portrayals aligned with participants' accurate belief that suspicious emails can be a threat vector.

## 5 Evaluating realism in fictional portrayals of computer security

Our third research question focuses on why users learn these particular concepts about security from fictional media. Users make judgments about realism and importance that determine whether and how these fictional portrayals are incorporated into their overall mental models. Understanding these judgments is compelling because, if we can identify how and why a misconception is formed, we may be better able to prevent it, or even to work within incorrect or incomplete mental models to nonetheless promote better security outcomes. To answer this question, we examined how participants assessed the accuracy of the clips we showed.

Participants exhibited a variety of heuristics for assessing the realism of computer-security incidents and behaviors in fictional media. In particular, most participants started from a default assumption about how likely such media are to be realistic. This default assumption was then mediated by other cues specific to a particular clip or scene. We categorize these additional cues into four major categories:

- *Technical knowledge*, which helps participants who understand some aspects of the depicted events evaluate how realistic they are;
- *Non-technical experience*, in which participants relate the depicted events to their own lives;
- *Plausibility of plot and characters*, in which participants consider whether the motivations and behaviors of characters, together with the broader events of the plot, are reasonable;
- and *Cinematic aspects*, in which visual and audio cues such as set decoration, musical score, and internal consistency affect the participants' evaluations of the scene.

We detail examples from each of these categories, plus default assumptions, in the subsections below.

## 5.1 Default assumptions about realism

Despite the variability across participants in opinions about the accuracy of the media, we noted that each individual participant's beliefs seemed to default to a particular attitude toward the media's accuracy: accurate, inaccurate, or mixed. When participants lacked sufficient cues to help them decide whether a clip was accurate, or had difficulty understanding the clip, they typically relied on this default opinion. For example, P16 — who said in the final phase of the interview that he found media portrayals generally accurate — also said of the Blackhat clip, "I really didn't get much from it, but I think it falls on the type of realistic thing."

These general views of media accuracy, and the resulting default assumptions about accuracy in individual clips, corresponded to specific opinions about the motivations of the media in presenting computer security information. The eight participants who said the fictional media was presenting a generally accurate picture of computer security tended to believe that one goal of these portrayals is to educate viewers. As P18 noted, "They're [entertainment media] doing a good job I believe, yeah. You learn from it."

Five participants said media generally present computer security unrealistically. Some attributed this to the creators' lack of expertise in the field. This was exemplified by P14: "I'm going to say they're portrayed inaccurately, because I don't think any of the people directing or creating these movies have really experienced being hacked, and I think that's why they're so dramatic and over the top and fast."

Others instead believed entertainment media were intentionally hiding information about computer security from viewers. These participants were unlikely to trust the portrayals we showed and likely to believe that all forms of security are futile (Section 4.3). For example, P6 said "I think it's [fictional media] probably inaccurate because we all know the government don't allow you to show everything that goes on. . . and they're not going to put realistic things for everyone to view." As mentioned above, P6 also found the portrayal of a master encryption key in Sneakers realistic.

Finally, six participants assumed that fictional portrayals were a mix of realistic and not, resulting from the creator's choice to trade off presenting security information accurately and providing a compelling plot. Accurate portrayals were seen as too boring to sell well. P7 describes this feeling: "I think it would be hard to portray the actual process so it would keep the flashy appeal to mass audiences that those movies target, because I imagine it's something that slowly develops over a long period of time and there's a lot of trial and error. And it's not just use a button and flashy things happen. It's like I imagine it would be more subtle."

Redmiles et al. note that uptake of security advice is commonly mediated by trust in the advice giver rather than evaluation of the advice content [21]. This reinforces our finding that people's default trust in fictional media will affect the way they process and absorb fictional depictions.

Participants' reliance on these default assumptions of accuracy, however, was mediated by a variety of cues and proxies that signal realism in a particular clip, as described in the following sections.

## 5.2 Technical knowledge

Participants frequently tried to use their pre-existing technical knowledge as a basis for discriminating between realistic and unrealistic depictions of computer security. However, participants often did not have enough technical knowledge to fully evaluate a clip, falling back instead on various proxies:

**Jargon typically implies technical realism.** Many characters in the selected clips demonstrate technical expertise by reeling off litanies of technical terms, and participants often responded to these unfamiliar terms by assuming the clips must be realistic. In particular, 12 participants assumed that if they don't understand what is being said, the person speaking must be knowledgeable, and thus their words and actions must be plausible. This was the case for P1, who referenced the jargon in the Gumball scene: "What she said was realistic . . . it's like a foreign language to my ears, like when it's a doctor and you have no knowledge of what they talk about." P10 agreed, saying "Some of the words she was saying, like proxy and all that, I was like oh my gosh she knew her stuff. Yeah that was realistic." But not every participant was so quick to trust explanations filled with obscure technical terms. P11, for example, believed jargon might indicate an unrealistic attempt to sound technical without real accuracy, noting that

"It may have been someone pulling a lot of technical terms and throwing them into a paragraph."

**If it's too fast or easy, it's not realistic.** Another common heuristic used by interviewees was to assess whether the level of difficulty portrayed in the scene seemed appropriate for the task, with 12 participants noting that surprisingly fast or easy tasks seemed unrealistic. P14, for example, commented on how easily the defenders in Skyfall noticed they were being hacked: "Even like once the hacker hacks you, I don't think it would be as easily identifiable." P12 drew a similar conclusion about the defenders in NCIS: "How sudden and fast it seems like, and the fact that she knew it was happening." Participants' tendencies to assume that computer security defense must be difficult and advanced, and be surprised when it seems too simple, may be related to findings by previous researchers that users find computer security advice advanced and intimidating, and feel helpless to take appropriate action to protect themselves [13, 22].

## 5.3 Non-technical background

Participants also used their non-technical background and experience, including the relatability of characters in a scene, to inform their realism judgments.

**If it matches a negative personal experience, it's realistic.** Eleven participants assessed realism by connecting on-screen events to a previous negative experience in their own life. This often led them to believe a scenario was more likely to be accurate. For example, in response to Blackhat, P13 said, "He used a keylogger to find out his password from a email and a download, which I believe is totally possible. I had a weird thing where my stepdad put a keylogger on my computer to see if I had a Facebook. So I know that this is possible." These findings fit with previous research identifying negative personal experiences as one important source of security mental models [21, 22].

**Relatable events are realistic.** When participants did not have relevant personal experience to draw from, they often used the relatability of a clip — whether or not they could imagine having the same events happen to them or behaving as the on-screen characters did — as a proxy for realism. In response to the phishing attack in Blackhat, P8 said "I probably would've fallen victim to it too. Anybody else would, it seems like a credible thing referencing an email the way it did." This was echoed by P9, who said of the NCIS clip: "It was almost like a rash, you couldn't do anything to stop the itching or the burning, it was moving so fast. That's the part I could identify with." This aligns with Redmiles et al.'s finding that negative experiences depicted in media can serve as a learning tool for security behavior when the characters are relatable [22], as well as Moyer-Gusé's theory that character identification can increase retention and behavior change with respect to educational entertainment more broadly [17].

## 5.4 Compliance with existing folk models

Many participants judged the overall realism of a clip in part by the extent to which it seemed to plausibly reflect their existing, non-technical beliefs about how the world works and how hackers behave. As discussed above, this can create a feedback effect whereby sufficiently common tropes influence users' beliefs and their judgments about subsequent media.

**Motivation for hacking matters.** Eleven participants noted that the attacker's motivation informed their overall evaluation. Watching Superman 3, multiple participants suggested that the main character's motivation to steal residual money from the pay system because he was disgruntled about the size of his paycheck was realistic. P19, for example, noted that "a lot of people feel like they are unpaid for the work that they do. I can relate: with various jobs, I felt like I was underpaid, and a lot of people feel the same way at their jobs."

Participants similarly believed that real-world hackers are motivated by the desire to flaunt their talent; thus, attackers trying to prove their talent or intelligence was taken as a signal of realism in a clip. P8, for example, found the Gumball clip believable because "She was showing off, and she enjoyed showing off. . . . If you're good at it there's inclination to want to be very good at it, to show 'em who you are." This mirrors Wash's folk model of hackers as "graffiti artists," motivated to attack to show off [27].

Relatedly, participants were skeptical of on-screen hackers' motivations when they did not believe the tradeoff between the cost and benefit for the hacker added up. P14 expressed mixed feelings about the realism of Superman 3: "Yes, as an individual you want to try and get your money, but no, because when it comes to the government the risks are so high that I don't think the cost-benefit is worth it."

**High-value targets imply realism.** As discussed in Section 4 above, our participants tend to believe that hackers exclusively target specific, high-value victims. This fed back into a belief that government targets and targets with high monetary value made a scene more realistic (n=5). In reference to the primarily governmental targets in Sneakers, P6 observed, "Well, the only other thing that I found realistic was the things that they were breaking. The codes were very high security." This, too, mirrors a mental model identified by Wash, that "hackers are criminals who target big fish" [27].

**Violating hacker stereotypes is not realistic.** Instead of focusing on the target, five participants zeroed in on the hacker, using their perceptions and stereotypes about who hackers are to evaluate the validity of the plot. When the portrayed hackers didn't match their expectations, they found the entire scene less plausible. For example, when asked about Gumball, P5

stated, "I don't think someone that age could do that." P8 said that Blackhat was unrealistic in part because of "how good the actor [Chris Hemsworth] looks, I guess." Here Wash's folk models play out in a more general way, with participants questioning characters who fail to fit into any of the hacker mental models they have available [27].

**Existence of consequences helps determine realism.** Participants also used the consequences for hackers to inform their overall evaluation. For example, two participants noted that the clips often included acts that were against the law, and that they therefore expected the attackers to be punished. P2 stated, "I think that when someone is hacking into government stuff or agencies, I feel like the authority will be alerted or someone or something will be alerted, and they will take action. I feel like in the clips they don't show, like, police or government taking any action, or someone alerting them that someone is hacking into their system," and that made the entire clip seem unrealistic.

In contrast, when authorities did intervene, some participants felt that this increased the realism. In reference to the NCIS clip, where the defenders actually are the authorities, P1 said that the clip was realistic because "the outcome of the show — they caught the guy, you know, how the whole thing played out." While this was not a common observation, some participants did use this as a proxy for realism.

**Hacking is plausible.** More broadly, participants considered whether the events of the plot were plausible, extrapolating from their beliefs about how the real world works. Fifteen participants, for example, mentioned that at least one clip was realistic in part because hacking does happen often in real life. In reference to Skyfall, P10 said that "companies do get hacked. That's how it's real." Occasionally, the likelihood of hacking in general was the only tangible thing participants could connect with reality, as demonstrated when P12 answered that the only realistic thing about the NCIS clip was "that it happens, people do get hacked."

**Repeated tropes are more realistic.** Another indicator of realism for participants was the popularity of plot points across the entertainment industry. Discussing Superman 3, two participants said they had previously seen other depictions of a disgruntled employee collecting fractions of cents shaved off of other employees' paychecks. Similarly, P10 was initially unsure about the realism of hacking of power grids in Sneakers, because he "never heard about it in real life happening, like someone taking over the city lights or whatever," but eventually concluded that the scene was realistic because he had "seen it in plenty of movies." This fits well with our overall finding that fictional tropes help to develop mental models, which are then reinforced by repeated exposure. It also aligns well with findings by Redmiles et al. that participants are more likely to trust information they are given based on "how widespread the advice was on various media outlets" [22].

## 5.5 Cinematic aspects

Finally, participants often cited aspects of the clips that were intrinsic to the medium of fictional television and film as a proxy for determining realism.

**Visual and audio cues affect realism.** One influential cue was the visual quality of a scene, which 17 participants pointed to when determining what was realistic and what was fantasy. Participants were split on whether overt demonstrations of so-called "Hollywood Hacking" — a common set of visual indicators that signify to an audience that hacking is occurring within the context of the movie, such as the red-skull calling card in the Skyfall clip — were realistic or not. Auditory effects also played a role. P8 noted of the Sneakers clip, "It reminded me actually of [Skyfall] because of the music escalating and being very overt and very dramatic and trying to move the plot along." For this participant, the dramatization made the scene feel less realistic.

**Physical and temporal setting have to fit.** Participants often used the set and setting of a scene to determine how realistic the scene was. Several participants called into question physical aspects of scenes that were incongruous with the rest of the environment, resulting in an overall judgment that a clip was unrealistic. In Skyfall, for example, a hacking attempt is portrayed as successful by showing several glass containers on the floor of an office building opening on their own. According to P16, "What looked less realistic was the tops coming out of the floors, I don't know what that was." P12 said, "I think it was kinda funny how they had showed the underside of the keyboard literally logging the keys," in response to the Blackhat clip, where the camera focuses on a seemingly unrelated portion of the scene for dramatic effect.

Temporal settings raised similar concerns. Two of our clips are decades old (Superman 3, Sneakers); three participants had a difficult time balancing whether they thought a task was realistic in the present, compared to in the time period the clip was portraying. According to P13, "I have no idea if that's realistic or not, because that is super old, so maybe it's realistic, he maybe could have done it, but . . . I don't know what the super low-level security would be at that time."

**Character behavior must be realistic.** The general behavior of characters, both targets and hackers, also impacted perceptions of realism, both positively and negatively. For example, P9 cited "the emotion, his reaction that he'd been duped," in Skyfall when judging it as a realistic scene, because the character responded as she would have expected. Similarly, in the NCIS clip, P3 pointed out that "the team supporting each other" seemed real. Interestingly, other participants used the same logic to judge the same scene as less realistic. For example, P11 thought the NCIS clip was unrealistic because "everybody just seemed too casual about it. A guy is eating a sandwich and saying what's going on, is this a video game?"

**Portrayals on screen need to match explanations.** "Show, don't tell" is a common piece of advice for writing, but this advice is not always followed. In the Gumball clip, the main character claims that she went through a lengthy process in order to hack into and open a door, but is only shown typing in four letters. While this discrepancy seems intentional, for comedic effect, three participants latched onto it to explain why they felt her actions were unrealistic. P15 commented: "I think what was kind of realistic was how she described in depth how she hacked it, you know, it just sounds complex and like there are probably some things they have to do to hack the system. But the unrealistic, you know, is that H-A-C-K hack." P6 explicitly noted that "She named so many things that she did and she only pushed 3 or 4 buttons. I mean, the only thing I saw her do was open the door." In the real world, a hacker or security professional could certainly kick off a large series of complex steps by issuing a single command, e.g. to run a script. However, without showing that laborious process, viewers are left to wonder how the simple action they saw relates to the complicated process that was discussed.

**Incongruity reduces realism.** Apparent randomness, or the generally incongruous nature of various elements in the clips we tested, also affected perceived realism. For example, seven participants expressed doubt about the realism of the Gumball clip just because it was a cartoon. P1 noted that the clip is not believable because "it's a cartoon." P18 agreed that "I don't think animation can be real," and P16 commented that "it's a cartoon, so you know, this is for kids." The technical jargon in this clip was so incongruous with the childlike nature of the visual elements that people perceived it as unrealistic, even though it was in some respects the most technically realistic scene showed during the interview. Even mundane plot elements that seemed out of place affected participants' interpretation of the clips. For example, P12 thought the conversation that led to the hacking in Superman 3 seemed forced, saying "I don't know why that coworker would randomly tell him about that."

## 6 Discussion

Our interviews demonstrate that users draw conclusions about what is (not) realistic about computer security in fictional media using a variety of heuristics, most of which are either entirely non-technical or only partially grounded in technical understanding. Further, many users believe that these media portrayals are either mostly, or at least partially, accurate: eight participants believed portrayals were generally accurate, six believed they were mixed, and only five concluded they were primarily inaccurate. This has important implications for users' mental models, as we know from previous studies that fictional media is one important factor in establishing these models for security behaviors [21, 23]. If the mechanisms users apply when deciding which information to adopt from fictional media are mostly divorced from even approximate technical correctness, and this media frequently presents unrealistic depictions, then users will be left with inaccurate and potentially harmful mental models.

Indeed, we see this play out in our study. Several clips were chosen because of their inaccuracies. Despite this, participants often failed to identify obviously unrealistic behavior. For example, it was common for participants to watch the Sneakers clip and conclude that widespread breaking of encryption is plausible and perhaps even occurs commonly in reality. While there are many existing vulnerabilities that place existing systems at risk, the belief that nothing can be safe, inculcated in part by television and film, can have negative consequences for users. This echoes work by Wash and Rader that found that users who believe that there is no way to make something secure often conclude that efforts to defend themselves and use good security behaviors are pointless [28]. Similar results, in the context of encrypted messaging, were observed by Abu-Salma et al. [2]. Even in cases of more benign errors, such as when two defenders in the NCIS clip worked together on the same keyboard to frantically defend against a hacker in real time, participants' consistent failure to notice the inaccuracy may be cause for alarm. Despite occasional success at pointing out other unrealistic aspects of the scene, the overall willingness to credit a scene with such an obvious inaccuracy, with one participant even noting the defenders working together as realistic, raises concern about the effect of inaccurate media on viewers.

**The need for collaboration.** Our findings point to the need for collaboration between the entertainment industry and the computer-security community. The entertainment industry has strong institutional knowledge in maintaining viewer engagement, but often seems to lack either the technical knowledge or the desire to depict security reasonably realistically, in a way that improves people's ability to make good security-relevant decisions. The academic security community, in contrast, has desirable lessons to teach users, but lacks a wide-scale platform to do so. One possibility for future work is to explore how to improve depictions of computer security in fictional media and evaluate how these improvements might affect users' understanding and decision making. We are particularly interested in the parallel to the field of medicine: In this field, the American Medical Association has issued guidance cracking down on pseudoscience and inaccuracies in the media, and medical advisors have been hired for films and television [4, 19]. Studies assessing the impact of these interventions on viewers have demonstrated positive impact almost three times as often as negative [12]. We are intrigued by the possibility that analogous interventions related to computer security could have a similar positive impact on viewers' knowledge and behavior.

To this end, we propose a Cybersecurity in Entertainment Task Force to mediate between the entertainment industry and the security community. Additionally, we encourage television and film productions that intend to portray computer security or hacking on-screen to hire technology advisors. These suggestions parallel the science and medical advisors many productions already hire, as well as the work done by the Hollywood, Health, and Society organization, which has worked with 91 TV shows to provide consultants and accurate medical information [1]. Further, there is good evidence that accurate portrayals of hacking can indeed be entertaining. For example, the television show Mr. Robot, which has been lauded for its accurate depictions of computer security, has also enjoyed critical acclaim [9, 31].

**Entertaining responsibly.** Although it is unclear whether users are learning from fictional media or fictional media is reinforcing their already existent mental models (or both), it is clear that media portrayals include known technical fallacies. Some of these inaccuracies matter more than others in terms of what viewers ultimately take away from scenes. If what viewers learn from an inaccurate scene is that two hackers can use one keyboard in an emergency, or that it only takes a moment to break into a secure headquarters, their own security behaviors are unlikely to be negatively affected. However, if they instead learn that all encryption is broken, that hacking is always obvious and easy to identify, or that the best way to respond to a breach is to restart your computer, they may make bad security decisions. By choosing which forms of inaccuracy to portray, creators of entertainment can still create exaggerated scenes filled with fast-paced action and sensationalism, while avoiding imparting particularly problematic misconceptions to their viewers. Further, our results identify heuristics that convey not only realism, but lack of realism. When presenting potentially harmfully inaccurate information, the media could provide cues not to take it seriously, mitigating the harm done. Further, even when it is not possible to convey all details accurately, ensuring that depictions of computer security are at least reasonable at a high level would still be a strong improvement.

**Guidance for researchers and educators.** Security researchers and educators, of course, may not be in a position to change the habits of entertainment producers. Our findings, however, also provide insight to help researchers and educators cope with misinformation disseminated in fictional media. Educators, researchers, and designers who better understand common tropes, and the misconceptions they lead to, can address these tropes directly in security tools, informational messages, and other guidance by pointing out explicitly what users may misunderstand. Alternatively, interventions could try to work within existing tropes by adapting advice, tools, or interfaces to fit existing mental models. Further, researchers and practitioners can take advantage of these tropes to lend realism and seriousness to their own informational messages and examples. Designing security interventions that will be perceived as realistic and relatable could help users understand and adopt better threat models and behaviors.

**The media is not a monolith.** In this paper, we explored fictional U.S. television and film about computer security. Future work could examine how cybersecurity is portrayed in other regions, in non-fiction and news, or in other types of fictional media, like books or podcasts. It might be particularly interesting to consider whether the specific media properties that users consume directly affect the mental models they end up with; however, untangling this possibility from other factors that might inform a user's mental model may prove challenging. A related question concerns genre in fictional media: do certain genres tend toward portrayals that do a better or worse job of developing accurate mental models in users, or do users who primarily watch particular genres develop more realistic mental models?

## 7 Conclusion

In this paper, we interviewed 19 participants about their mental models of computer security, hacking, and encryption, and how those mental models were influenced by portrayals of these concepts in fictional media. To focus on the role of media in forming mental models, we showed interviewees six clips from television series and films depicting computer-security topics. We asked participants what they considered realistic and why, in these individual clips and in fictional-media depictions of computer security as a whole.

We find that people incorporate fictional portrayals into their mental models of computer security, with sometimes unfortunate effects. Participants typically used proxies, many of which were non-technical, to evaluate the accuracy of particular depictions of computer security. Further, these models — in part drawn from popular depictions — can be self-reinforcing, as additional exposure to common tropes serves to confirm participants' pre-existing beliefs.

We therefore conclude that media portrayals of computer security contribute to the development of incomplete and inaccurate mental models. So long as this remains true, common fictional tropes must be taken into account when seeking to improve security education. To address this challenge, we propose a closer partnership between the computer-security field and the entertainment industry, we suggest approaches for the entertainment industry to provide entertainment while avoiding inculcating misconceptions, and we recommend that security researchers and educators take the effects of fictional portrayals into account when trying to teach users about security concepts and behaviors.

## References


[1] History: HH&S by the numbers. https://hollywoodhealthandsociety.org/about-us/history-hhs-numbers.

[2] R. Abu-Salma, M. A. Sasse, J. Bonneau, A. Danilova, A. Naiakshina, and M. Smith. Obstacles to the adoption of secure communication tools. In *2017 IEEE Symposium on Security and Privacy*, pages 137–153, May 2017.

[3] Alessandro Acquisti and Jens Grossklags. Privacy and rationality in individual decision making. 3(1):26–33, 2005.

[4] Julia Belluz. The American Medical Association is finally taking a stand on quacks like Dr. Oz. https://www.vox.com/2015/6/13/8773695/AMA-dr-oz#, Jun 2015.

[5] Alain Forget, Sarah Pearman, Jeremy Thomas, Alessandro Acquisti, Nicolas Christin, Lorrie Faith Cranor, Serge Egelman, Marian Harbach, and Rahul Telang. Do or do not, there is no try: user engagement may not improve security outcomes. In *2016 Symposium on Usable Privacy and Security*, pages 97–111, 2016.

[6] Deen G Freelon. ReCal: Intercoder reliability calculation as a web service. *International Journal of Internet Science*, 5(1):20–33, 2010.

[7] Damian Gordon. Forty years of movie hacking: considering the potential implications of the popular media representation of computer hackers from 1968 to 2008. *International Journal of Internet Technology and Secured Transactions*, 2(1/2):59–87, 2010.

[8] Greg Guest, Arwen Bunce, and Laura Johnson. How many interviews are enough? an experiment with data saturation and variability. *Field Methods*, 18(1):59–82, 2006.

[9] Aisha Harris. The fourth season of 'Mr. Robot' will be its last. https://www.nytimes.com/2018/08/29/arts/television/mr-robot-last-season.html, Aug 2018.

[10] Andrew F Hayes and Klaus Krippendorff. Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, 1(1):77–89, 2007.

[11] Cormac Herley. So long, and no thanks for the externalities: the rational rejection of security advice by users. In *2009 Workshop on New Security Paradigms Workshop*, pages 133–144. ACM, 2009.

[12] Beth L. Hoffman, Ariel Shensa, Charles Wessel, Robert Hoffman, and Brian A. Primack. Exposure to fictional medical television and health: a systematic review. *Health Education Research*, 32(2):107–123, 2017. http://dx.doi.org/10.1093/her/cyx034.

[13] Ruogu Kang, Laura Dabbish, Nathaniel Fruchter, and Sara Kiesler. "My data just goes everywhere:" user mental models of the internet and implications for privacy and security. In *2015 Symposium on Usable Privacy and Security*, pages 39–52. USENIX Association Berkeley, CA, 2015.

[14] Klaus Krippendorff. Reliability in content analysis : Some common misconceptions and recommendations. 2015.

[15] Matthew Lombard, Jennifer Snyder-Duch, and Cheryl Campanella Bracken. Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research*, 28(4):587–604, 2002.

[16] Johnny Long. Secrets of the hollywood hacker! In *DefCon 14*, 2006.

[17] E. Moyer-Gusé. Toward a theory of entertainment persuasion: Explaining the persuasive effects of entertainment-education messages. *Communication Theory*, 18(3):407–425, 2008. http://dx.doi.org/10.1111/J.1468-2885.2008.00328.X.

[18] Delroy L. Paulhus. Chapter 2 - Measurement and control of response bias. In John P. Robinson, Phillip R. Shaver, and Lawrence S. Wrightsman, editors, *Measures of Personality and Social Psychological Attitudes*, pages 17 – 59. Academic Press, 1991. http://www.sciencedirect.com/science/article/pii/B978012590241050006X.

[19] McKenna Princing. I was a medical advisor for Grey's Anatomy. Here's what i learned. https://rightasrain.uwmedicine.org/well/stories/i-was-medical-advisor-greys-anatomy-heres-what-i-le
2018.

[20] Emilee Rader, Rick Wash, and Brandon Brooks. Stories as informal lessons about security. In *2012 Symposium on Usable Privacy and Security*, page 6. ACM, 2012.

[21] Elissa M Redmiles, Sean Kross, and Michelle L Mazurek. How i learned to be secure: a census-representative survey of security advice sources and behavior. In *2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 666–677. ACM, 2016.

[22] Elissa M Redmiles, Amelia R Malone, and Michelle L Mazurek. I think they're trying to tell me something: Advice sources and selection for digital security. In *2016 IEEE Symposium on Security and Privacy*, pages 272–288. IEEE, 2016.

[23] Scott Ruoti, Tyler Monson, Justin Wu, Daniel Zappala, and Kent Seamons. Weighing context and trade-offs: How suburban adults selected their online security posture. In *2017 Symposium on Usable Privacy and Security*, pages 211–228. USENIX Association, 2017.

[24] Anselm Strauss, Juliet Corbin, et al. *Basics of qualitative research*, volume 15. Newbury Park, CA: Sage, 1990.

[25] Blase Ur, Jonathan Bees, Sean M Segreti, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. Do users' perceptions of password security match reality? In *2016 CHI Conference on Human Factors in Computing Systems*, pages 3748–3760. ACM, 2016.

[26] Reddit: /r/itsaunixsystem/. https://www.reddit.com/r/itsaunixsystem/, 2018.

[27] Rick Wash. Folk models of home computer security. In *2010 Symposium on Usable Privacy and Security*, page 11. ACM, 2010.

[28] Rick Wash and Emilee J Rader. Too much knowledge? security beliefs and protective behaviors among united states internet users. In *2015 Symposium on Usable Privacy and Security*, pages 309–325, 2015.

[29] David Knapp Whittier, May G Kennedy, Janet S St. Lawrence, Salvatore Seeley, and Vicki Beck. Embedding health messages into entertainment television: Effect on gay men's response to a syphilis outbreak. *Journal of Health Communication*, 10(3):251–259, 2005.

[30] Justin Wu and Daniel Zappala. When is a tree really a truck? exploring mental models of encryption. In *2018 Symposium on Usable Privacy and Security*. USENIX Association, 2018.

[31] Kim Zetter. How the real hackers behind Mr. Robot get it so right. https://www.wired.com/2016/07/real-hackers-behind-mr-robot-get-right/, Jul 2016.

## A Recruitment survey

**Please specify the gender with which you most closely identify.**

- Male
- Female
- Other
- Prefer not to answer

**Please specify your age.**

- 18-29
- 30-39
- 40-49
- 50-59
- 60-69
- Over 70

**Please specify your ethnicity.**

- White
- Hispanic or Latino
- Black or African American
- American Indian or Alaska Native
- Asian, Native Hawaiian, or Pacific Islander
- Other

**Please specify the highest degree or level of school you have completed**

- Some high school credit, no diploma or equivalent
- High school graduate, diploma or the equivalent (for example: GED)
- Some college credit, no degree
- Trade/technical/vocational training
- Associate degree
- Bachelor's degree
- Master's degree
- Professional degree
- Doctorate degree

**If you are currently a student or have completed a college degree, please specify your field(s) of study (e.g. Biology, Computer Science, etc).**

- Text field

**Please select the response option that best describes your current employment status.**

- Working for payment or profit
- Unemployed
- Looking after home/family
- A student
- Retired
- Unable to work due to permanent sickness or disability
- Other [text field]

**If you are currently working for payment, please specify your current job title.**

- Text field

**Please write in the number of hours you typically spend on each of the following activities in the specified time range:**

- recreational TV: ___ hours /week
- newspaper: ___ hours /week
- podcasts: ___ hours /week
- social media: ___ hours /day
- movies: ___ hours /month
- TV news: ___ hours /week
- magazines: ___ hours /week

**Which genres do you enjoy consuming media in (select as many as you want):**

- action
- comedy
- romantic
- documentary
- horror
- drama
- kids
- adventure
- sci fi
- fantasy
- other
- thriller/spy films

**Please enter your email address so the we can contact you for the interview, if you are selected.**

**Your contact information will only be used to invite you to participate in the study. After the study, all records of your contact information will be destroyed unless you indicated above that you agree to be contacted regarding future studies.**

- Text field

## B Interview protocol

**Introduction**

- Hello. My name is [name] and this is [name]. Today we will be conducting a study to learn about what you may heard about cyber security threats.
- First, let's quickly go over how the study will work. We will break the session into three parts: questions about how you perceive cyber threats, the presentation of some short audio visual clips, and questions about your reaction to those clips. I expect the study will take approximately one hour. This is not a quiz or test of your knowledge; in fact there are no correct answers. We only want to learn about what you have heard about cyber security threats.
- Describe everything in the consent form.
- Although I do not expect this to occur, if you become uncomfortable at any time during the study please let me know. Do you have any questions at this point?
- Give the subject the consent form. I have this consent form here. It tells you whom to contact if you want to report any objections. I'll give you two copies - one is for you to keep, and the other is for you to sign.
- Ppint out places the subject needs to sign; point out the section where it states they will be auditorily recorded.

**Phrase association and mental models**

- To begin I'm going to ask you for any associations you have with some buzzwords. What comes to your mind when you hear:
  - Cybersecurity
  - Hacker
  - Encryption
- Now I'm going to ask some specific questions about your beliefs with regard to cyber security topics. Remember there are no wrong answers; I'm curious about your perceptions.
  - Can you describe to me what a hacker's goals are in general?
  - What makes someone an easy target for a hacker?
  - Can you describe to me who a hacker's intended target normally is?
  - What are some ways that users and businesses implement cyber security?
  - What, if any, limitations to hackers have? What mechanisms do they hackers utilize?
  - What is encryption used for?
  - How do you think people become involved with the hacking community?
  - What are some ways people can defend themselves against hackers?
- Where do you think those ideas come from? That is, what influences your perception of those ideas and phrases?

**Personal experiences**

- Now let's talk about your experiences with these topics. Have you or someone you know been hacked?
  - How did that happen/ What do you think happened?
  - What alerted you/them to the fact that their security had been breached?
  - What steps did you/they take to remedy the issue?

**Prior media exposure**

- Have you ever seen any depictions of the terms or ideas we've been discussing in fictional media? In a fictional TV show or in a movie? For example of a hacker?
  - Where?
  - How recent is/was it?
- Can you name any specific fictional TV shows or movies that have such depictions?
  - Can you think of any specific examples of a scene about [term] that you saw recently? Can you explain what was depicted?

**Clip presentation and reactions**

- Now we're going to move to the second section of the interview. I am going to show a series of short clips from various fictional movies and T.V shows, and then ask some questions about what you've seen immediately after each one. Before we begin, do you have any questions about what we've discussed so far?
- Play a clip, with the video in full screen mode.
  - Have you seen this show/movie before?
    - If so, have you seen this specific scene or clip?
  - Can you give me an overview of what you think is happening in the scene?
  - Scenes from fictional TV/movies often have some aspects that are realistic and some that are less realistic. What did you think was realistic about this scene?
    - Why do you find that aspect realistic?
  - What did you think was unrealistic about this scene?
    - Why do you find that aspect unrealistic?
- Repeat this same process with five additional clips.

**Post-clip responses**

- We're now beginning the final stage of the interview. I'm going to ask some questions about media portrayals in general. But first, do you have any questions about what we've covered so far?
- Do you feel the media, specifically fictional TV and movies, portrays cybersecurity, hackers, and encryption accurately, and why?
- Can you think of any fictional TV shows/movies that portray the topics we've discussed today realistically?
- Conversely, can you think of any fictional TV shows/movies that portray the topics we've discussed today unrealistically?

**Closing**

- That brings us to the conclusion of this interview! Do you have any final thoughts or questions?
- Thank you so much for your time. Here is your compensation and the consent form for your records.
- Give participant compensation and have them sign a receipt that they were paid. Also give them the unsigned consent form for their records.

**Clips shown**

- Skyfall: https://www.youtube.com/watch?v=aApTVqeGJMw
- Superman 3: https://www.youtube.com/watch?v=iLw9OBV7HYA
- Sneakers: https://www.youtube.com/watch?v=F5bAa6gFvLs
- NCIS: https://www.youtube.com/watch?v=u8qgehH3kEQ
- The Amazing World of Gumball: https://www.youtube.com/watch?v=-rQPdWwv3k8
- Blackhat: https://www.youtube.com/watch?v=7HWfwLBqSQ4

# A Typology of Perceived Triggers for End-User Security and Privacy Behaviors

Sauvik Das
*Georgia Institute of Technology*
sauvik@gatech.edu

Laura A. Dabbish
*Carnegie Mellon University*
dabbish@cs.cmu.edu

Jason I. Hong
*Carnegie Mellon University*
jasonh@cs.cmu.edu

## Abstract

What triggers end-user security and privacy (S&P) behaviors? How do those triggers vary across individuals? When and how do people share their S&P behavior changes? Prior work, in usable security and persuasive design, suggests that answering these questions is critical if we are to design systems that encourage pro-S&P behaviors. Accordingly, we asked 852 online survey respondents about their most recent S&P behaviors ($n = 1947$), what led up to those behaviors, and if they shared those behaviors. We found that social "triggers", where people interacted with or observed others, were most common, followed by proactive triggers, where people acted absent of an external stimulus, and lastly by forced triggers, where people were forced to act. People from different age groups, nationalities, and levels of security behavioral intention (SBI) all varied in which triggers were dominant. Most importantly, people with low-to-medium SBI most commonly reported social triggers. Furthermore, participants were four times more likely to share their behavior changes with others when they, themselves, reported a social trigger.

## 1 Introduction

A longstanding goal in usable security and privacy is to bridge the gap between behaviors that experts recommend (pro-S&P behaviors) and those that end-users actually adopt [9, 18, 29]. However, these pro-S&P behaviors remain rare. For example, as of early 2018, fewer than 10% of Google account holders had enrolled in two-factor authentication and at least 17% of Google users re-used their account passwords [33].


*USENIX Symposium on Usable Privacy and Security (SOUPS)* 2019.
August 11–13, 2019, Santa Clara, CA, USA.

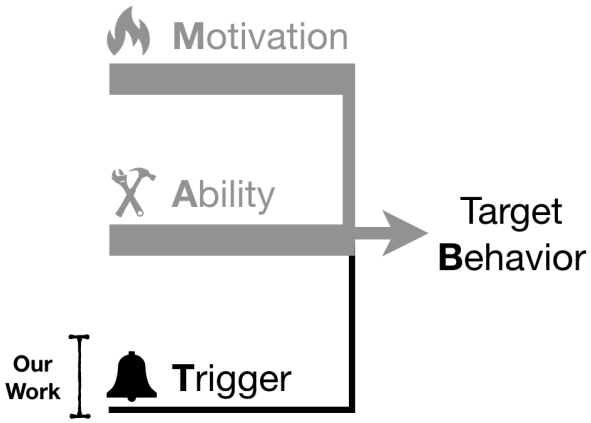



Figure 1: A popular model in behavioral psychology suggests that human behavior is a product of motivation, ability and trigger [20]. Prior work has extensively documented end-user motivation and ability to adopt S&P behaviors, but has less to say about behavioral *triggers* in-the-wild. We begin to systematically typify these behavioral triggers that lead to behavior change in S&P.

Recent Pew surveys found that only 12% of Internet users in the U.S. use password managers [36] and only 22% of smartphone users both use screen locks and regularly update their phone [4]. Additionally, Ion, Reeder and Consolvo showed that the pro-S&P behaviors that experts recommend only thinly overlap with the behaviors that non-experts find important and adopt [29].

The first step in bridging this disconnect between what experts *recommend* and what end-users *practice* is to understand what triggers pro-S&P behaviors when they *do* occur. Drawing from literature in behavioral psychology and persuasive design, a popular model suggests that behavior is a product of motivation, ability and trigger (also referred to as *prompt*) [20], as shown in Figure 1. That is, people perform

a behavior when they want to (motivation), believe they can (ability), and feel like they should *at that moment* (trigger). This framing helps identify gaps in our existing understanding of end-user S&P behavior. While there is extensive prior work that systematically typifies factors that inhibit people's *motivation* and *ability* to adopt S&P behaviors (e.g., they are difficult [47], time consuming [24], not relevant to the task at hand [15]), there has been comparatively less work typifying the *triggers* that prompt S&P behaviors, in general.

Prior qualitative work has, however, identified both observed and hypothesized S&P behavioral triggers (e.g., [9, 10, 37, 40]). Surveying this prior work, we synthesized a set of three broad trigger types that often precede S&P behaviors—social, forced, and proactive. *Social* triggers are direct social interactions that prompt behavior change, e.g., a friend providing advice or observing others' security behaviors. *Forced* triggers are non-social, and capture external stimuli or situations that necessitate behavior change outside of the end-users' own volition: e.g., experiencing a personal data breach or an employer requiring one to update one's passwords regularly. Finally, *proactive* triggers are also non-social, and capture internal processes that lead to volitional or goal-oriented behavior change: e.g., unprompted one-off decisions to enable a screen lock or routine password updates. Note that these trigger types primarily capture S&P behavioral triggers that are *perceptible in the moments leading up to behavior*. There are other "triggers" that could influence end-user S&P behaviors over longer periods of time that are harder to perceive in the moment: e.g., social norms or cultural attitudes towards openness, transparency and privacy. In this work, we primarily focus on the former as our goal is to provide actionable insights for researchers and designers.

While prior work helps categorize these in-the-moment behavioral triggers, what's missing is a generalization of these qualitative findings to a broader sample. Indeed, what are the relative frequencies of these trigger types? How do they differently manifest for individuals from different demographic backgrounds and attitudes towards S&P? How often people actually share their S&P behaviors with others? Answering these questions is important if we are to design effective interventions that encourage pro-S&P behaviors — particularly for users who have low-to-medium security behavioral intention (SBI), or intention to behave in a manner consistent with expert-recommended security and privacy advice [18]. Moreover, while prior work has found that social triggers were particularly promising in motivating S&P behaviors for non-experts [9, 10], it remains unclear how often and for what behaviors different people encounter social triggers.

Our primary contribution is to address these gaps in the literature through an online survey ($n = 852$) in which we asked participants to report on recent S&P behaviors and what led to those behaviors. Through this process, we aim to address the following research questions:

- **RQ1**: How relatively frequent are the social, forced and proactive triggers that lead to S&P behaviors?
- **RQ2**: How does the relative frequency of social, forced and proactive triggers for S&P behaviors differ across people from different demographic backgrounds and levels of SBI?
- **RQ3**: How often and why do people share their S&P behaviors with others, and what factors correlate with this sharing?

Overall, we found that social triggers were most numerous—39% of all reported triggers were social, compared to 34% proactive and 26% forced. However, this trigger distribution varied significantly across different types of behaviors and individuals from different age groups, nationalities and levels of security behavioral intention. Perhaps most importantly, *people with low-to-medium SBI were far more likely to report changing their S&P behaviors in response to a social trigger*. Conversely, people with higher SBI were far more likely to report updating their behaviors proactively. We also found that participants were four times more likely to share their behaviors with others when their own behavior was preceded by a social trigger. In sum, our findings offer a unique new perspective in explaining end-user SP behaviors which opens up several promising new threads of research for designing tools that encourage pro-S&P behaviors.

## 2 Related Work

A popular model of human behavior, the Fogg Behavior Model (FBM), provides a helpful framing for understanding how to encourage pro-S&P behaviors. In brief, behavior occurs if and only if one *wants* to adopt the behavior (motivation), is easily *able* to adopt the behavior (ability) and something *prompts action* (trigger) [20]. We divide our survey of related work in usable privacy and security as they relate to these three categories of the FBM.

### 2.1 Ability

A broad survey of the usable privacy and security literature suggests that there are at least two barriers that reduce people's *ability* to adopt S&P behaviors: awareness and knowledge. First, many users lack awareness of relevant security threats and what can be done to protect themselves from those threats. For example, prior studies have found that insufficient awareness of security issues resulted in people constructing their own, often incorrect, model of security threats [2, 16, 46]. Stanton et al. found that a lack of awareness of basic security principles influenced a number of security mistakes, such as using a social security number as a password [44]. In an analysis of expert and non-expert users, Ion, Reeder and

Consolvo [29] found that non-experts were unaware of the strategies experts employed to protect themselves.

Security tools are also often too complex for end-users to operate [2, 47]. Indeed, for many security and privacy systems, there is a wide gulf of execution [35] between what users *want* and *know how* to do. For example, many users cannot distinguish legitimate versus fraudulent URLs, nor forged versus legitimate email headers [13]. Another study revealed how security features in Windows XP, Internet Explorer, Outlook Express, and Word applications are difficult for users to understand and utilize [21]. Wash found that many people hold "folk models" of computer security that are often incorrect, which leads to ignoring security advice [46]. More recently, the Pew Internet Research center found that the majority of Internet users have strong misconceptions about basic cybersecurity concepts [36].

## 2.2 Motivation

In addition to being unable, people may also simply not want to follow recommended security advice [9]. This lack of motivation can be attributed to a number of key psychological principles. First, stringent security measures are often antagonistic towards the specific goal of the end user at any given moment [15, 41]. For example, strong e-mail account security (e.g., using two-factor authentication), can delay a user access to her email for an intolerable amount of time [17]. Thus, users may reject SP advice when they expect or experience it to be too time-consuming or require too much effort [2, 15, 28, 41].

Furthermore, many people may understand security threats in the abstract but may not believe they, themselves, are at risk [2,46]. Herley argues that this perspective may be rational, as the expected monetized cost of a lifetime of following commonly recommended security advice (e.g., reading suspcious URLs) may be orders of magnitude higher than the expected monetized loss a compromised account [25]. Furthermore, while the benefits of security features are abstract and delayed (i.e., protection from a potential threat sometime in the future), the costs are immediate and concrete (i.e., additional time or effort now and forevermore) [1]. Indeed, security claims are often unfalsifiable — irrespective of present behavior, there is no guarantee of future security [26]. Furthermore, prior work has found that there may be a social stigma associated with use of expert recommended security tools and advice that further lowers people's motivation to be secure [9, 11, 22].

## 2.3 Triggers

While prior work provides a rich foundation for understanding motivation and ability in the context of end-user S&P behaviors, we have less understanding of the triggers that prompt S&P behaviors and how they vary in frequency and effectiveness across individuals. This is not to say that there is no work on behavioral triggers in S&P. For example, there is much work on improving adherence to and compliance with security warnings, as these warnings are commonly ignored and begrudged [3, 5, 10, 16, 19]. Similarly, there is a wealth of information about how to design privacy notices for different use-cases and scenarios (for a review, see [42]. Much of the research on security warnings and privacy notifications is centered around urging end-users to react to an external prompt — what we call "forced" triggers.

In a recent qualitative study, Das et al. introduced a typology of *social* triggers for S&P behaviors [9]. They found that nearly 50% of all reported S&P behavior changes were the result of an implicit or explicit social interaction with others [9]. Among these social triggers were "observing others", "sharing access with others" and "receiving advice from others." Others have also noted the broad efficacy of social triggers for S&P behavior change. Both Rader and Redmiles et al. found that informal word-of-mouth stories were effective social triggers for S&P behaviors [37, 40].

Prior work has also shown that people can sometimes be *proactive* in their S&P behaviors. For example, a series of studies found that people can be proactive in self-censoring content or otherwise adjusting their social media privacy settings to avoid later regrets [8, 43, 45].

There is also evidence that the efficacy and frequency of different S&P triggers might vary across individuals. Redmiles et al. [38] found that people from different socioeconomic backgrounds may respond to differently to advice received from different sources (e.g., notices from the workplace vs. informal stories from friends). There is also a growing body of evidence documenting how people from different nationalities and cultural contexts can have different S&P attitudes and behaviors [30, 32].

To date, most of our knowledge of S&P behavioral triggers, in general, is piecemeal — assembled together from an ensemble of studies that either implicitly note the presence of these triggers or that more thoroughly study a specific trigger. To our knowledge, we are the first to quantitatively explore and synthesize S&P behavioral triggers generally. In doing so, we provide insights on: (i) the relative frequency of in-the-moment, perceived triggers that inspire S&P behavior change; (ii) how those triggers might vary across different S&P behaviors; and, (iii) how different individuals respond to different triggers. Armed with this understanding, we make empirically grounded suggestions on how to effectively design behavioral triggers that encourage pro-S&P behaviors.

# 3 Method

We conducted an online survey on the Amazon Mechanical Turk (AMT) platform [1]. We selected AMT partially because of the ease of recruiting a large sample on the platform, and

[1] https://mturk.com

partially because the biases of AMT samples are well studied [23,30,39]. To ensure high-quality responses, we included two attention-check questions, or questions for which participants are given specific instructions on how to answer to gauge if they are carefully reading questions [23]. We only discuss participants who passed these attention checks. The specific questions we asked in our survey is provided in Appendix A. For brevity, we provide a high-level overview of the questions.

**Behavior change questions**: We started by asking participants which, if any, of the following four behaviors they did in the past 6 months. The behaviors we selected were:

- *Mobile Auth*: enabling or changing one's method of authenticating into a mobile device (e.g., smartphone, laptop, tablet or other portable electronic device);
- *App Uninstallation*: uninstalling a smartphone application, specifically for privacy or security reasons;
- *Password Updates*: changing or updating a password for an online account; and,
- *Facebook Privacy*: updating one's Facebook account privacy settings.

We selected *these* behaviors because they were chosen in the closest related prior work [9], a qualitative exploration that found that nearly 50% of all reported behavioral triggers for the aforementioned behaviors were social. We wanted to compare our own results to that benchmark. Given that the selected behaviors still represent a diverse subset of S&P behaviors, we believe that our results should generalize as well as any other subset of S&P behaviors.

If participants had not recently done any of the aforementioned behaviors, they were allowed to manually specify a different S&P behavior they recalled doing in the past 6 months. They could answer remaining questions in reference to this "other" behavior.

**Trigger questions**: For each S&P behavior participants recalled having done in the past 6 months, we next asked participants which, if any, of a set of behavioral triggers preceded their decision to perform the behavior. The options we presented were synthesized from a survey of related work. Participants were also able to manually write-in a different trigger if the provided options were insufficient.

We categorized each of these triggers into three higher-level categories that we synthesized from a reading of prior work and a discussion among the authors — social, forced and proactive. Social triggers are those that involve a direct social interaction either with somebody the participant knew personally or with whom the participant could observe and/or interact (i.e., experiencing a security breach from someone who the participant knew, lending one's device to someone else, observing others around them, or receiving advice). Forced triggers are non-social, and suggest the presence of an external catalyst that the participant did not specifically seek or desire (i.e., a warning dialog, an organizational policy, experiencing a personal data breach from a stranger). Finally, proactive triggers, while also non-social, involve conative processes internal to the participant or voluntarily seeking out information that directly leads to behavior change (i.e., habit or routine, no specific stimulus, reading news articles, actively looking through device settings or options).

We note that this higher-level taxonomy may have blurred boundaries: some triggers, like changing one's PIN due to lending one's device to a friend, could be considered either "social" or "proactive". However, we categorized each individual trigger into one higher-level category using the following well-defined process. First, if the trigger reflected any direct social influence or interaction, we categorized it as "social", even if it might also be "forced" or "proactive." We make this distinction based on findings from prior work which suggest that social triggers are uniquely motivating [9]. If the trigger did *not* reflect a direct social process, we categorized it as "forced" if the behavior change was either mandatory or forced by circumstance. Otherwise, we categorized it as "proactive," as a non-social, non-forced trigger suggests that participants made the change voluntarily either because of routine, because of personal preference, or because they actively sought out information.

Table 1 shows a list of all trigger options we presented, their mapping to the higher-level categories of social, forced and proactive, as well as the overall percentage of participants who reported having experienced the trigger prior to enacting the behavior. Table 2 shows the distribution of the higher-level trigger types, both overall and across individual behaviors.

**Social context questions**: For each of the social triggers participants selected, we asked additional questions to uncover the social context of those triggers. For example, if participants selected the "Received Advice" trigger, they were asked to specify their relationship with the person from whom they received the advice: friend, family member, significant other, colleague or other. If participants selected other, they were allowed to manually write-in a description of their relationship with that person. Participants who did not select any social trigger would not see any of these questions.

**Sharing questions**: We also asked participants if they shared their behavior with others. If they did share, we asked them to specify with whom (e.g., friend, family member, etc.) and how they shared the change (e.g., through face to face conversation, phone call, SMS or email). We then asked participants why they shared the change, giving them a range of options largely derived from prior work [9]. Examples reasons include "I noticed they were being insecure" and "I felt obligated to protect them". For brevity, we omit the complete list of responses here, but list them in Appendix A. We also discuss participants' rationale to share (Table 5.3) and not share in more detail in our results. Participants were, again, allowed to manually write-in an answer if none of the provided options were sufficient.

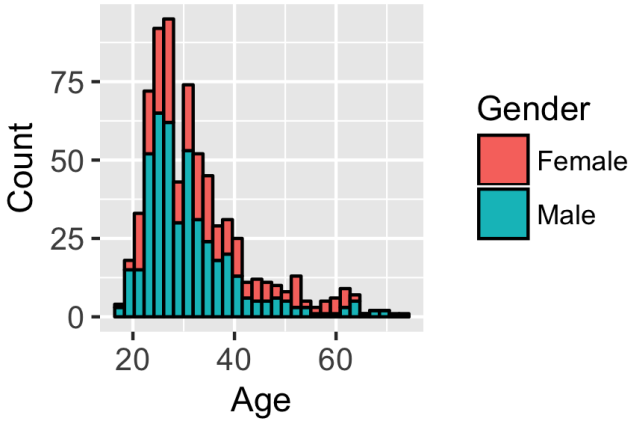



Figure 2: Distribution of participant ages and genders. Our participants were 33 years old, on average, and 63% self-reported as male.

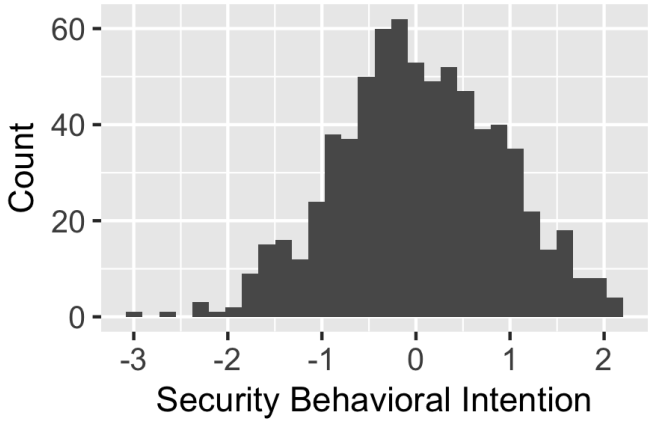



Figure 3: Distribution of security behavioral intention across all participants.

**Security behavioral intention questions**: We next asked participants to answer Egelman and Peer's SeBIS questionnaire [18] to measure their security behavioral intention (SBI). In addition, we asked participants a number of other questions about their general security knowledge and computer literacy derived from prior work.

**Demographic questions**: Finally, we asked participants to self-report a number of demographic dimensions, such as their age, gender and nationality. They were allowed to opt-out of providing any of this information, but nearly all participants answered all of the demographic questions.

### 3.1 Ethics and Compensation

Prior to data collection, we had our study approved by the Carnegie Mellon University Institutional Review Board (IRB). The survey took participants about 20 minutes to complete on average, and we paid participants $3.50 (translating to an hourly wage of $10.50). All collected data was anonymized — no identifiers were collected, and payment was facilitated through AMT.

## 4 Sample

Overall, we received responses from 1070 participants, 852 of whom both passed the attention-check quality tests and completed the entire questionnaire. Accordingly, we were left with $n = 852$ high-quality, complete responses.

### 4.1 Demographics and behavioral intention

Our participants had a mean age of $\mu = 33$ (sd: 10), ranging from 18 to 74. In addition, 537 participants self-reported as male (63%), 312 as female (36%) and 3 preferred not to answer. Figure 2 shows participants' age and gender distributions. Given the constraints of the AMT platform, our participants came mostly from the United States (449) and India (323). Approximately 47% of our participants reported a primary occupation that was "Computer Science related" or "Other engineering or technology related". However, only 10 reported occupations directly related to cybersecurity. Finally, 96% of participants reported being native English speakers.

To facilitate later analysis, we used a factor analysis to reduce the dimensionality of the 16-item SeBIS questionnaire [18] into one higher level construct we refer to as "security behavioral intention" (SBI). This single factor captured 17% of the variance in the responses to the SeBIS questionnaire, with each item being correlated with the construct in the expected direction (i.e., positively coded questions positively correlated, negatively coded questions negatively correlated). Figure 3 shows a histogram of the SBI in our population. Note that this method of dimensionality reduction has been previously used to facilitate analysis with the SeBIS [12].

### 4.2 Raw response counts for behaviors and triggers

**Behaviors**: Out of our 852 participants, in the 6 months preceding the survey, 454 (53%) reported having changed their mobile / laptop authentication settings, 427 (50%) reported changing their Facebook privacy settings, 378 (44%) reported uninstalling a mobile application for security and privacy reasons, and 688 (81%) reported changing a password for an online account or device. Overall, 807 participants (95%) reported doing at least one of the aforementioned four S&P behaviors and reported on 1947 behaviors, in total.

**Triggers**: Table 1 shows the distribution of triggers se-

lected for different behaviors. Across the 1947 behavior changes in our dataset, participants reported 2954 triggers leading up to those changes. A large majority of the triggers that lead to a security or privacy behavior were covered by the options we provided, which were based off a survey of prior work. Indeed, only 2% of reported triggers, overall, warranted manual specification that was not covered by our initial typology. Upon deeper investigation of these manual entries, most correlated strongly with the available triggers choices. The most commonly specified trigger that was not well covered by the questionnaire responses was "habit / routine" — a number of participants reported periodically, habitually or routinely updating their passwords, browsing their Facebook privacy settings, etc. We considered habitual / routine updating of S&P behaviors to be a proactive trigger.

## 5 Results

Recall that we had three high-level research questions we wanted to answer: (RQ1) How relatively frequent are social, forced and proactive triggers for S&P behaviors? (RQ2) How does the relative frequency of social, forced and proactive triggers for S&P behaviors differ across people from different demographic backgrounds and levels of SBI? and (RQ3) How often and why do people share their S&P behaviors with others, and what factors correlate with sharing? We present empirical answers to each of these questions.

### 5.1 RQ1: Relative trigger frequency

Table 2 shows the relative trigger frequency across all four behaviors. Overall, 1153 (39%) of reported triggers were social, 1005 (34%) were proactive and 773 (26%) were forced. To test if these rate differences were significant, we ran a logistic regression correlating trigger presence with trigger type (social, forced, proactive) and included random-intercept terms for distinct users and distinct behaviors to account for repeated observations (i.e., multiple behaviors per user, multiple triggers per behavior). We found that each of the pairwise rate differences were statistically significant ($p < 0.001$).

While different S&P behaviors vary in how often they are prompted by social, forced and proactive triggers, social triggers were the most frequent overall. This result highlights the importance of understanding and leveraging social influence to encourage better S&P behaviors. Indeed, out of all three of the higher-level trigger types, social triggers may be the most actionable. While many usable security interventions have attempted to make people more proactive about their security and privacy to little avail, the design space for encouraging greater social interaction in security and privacy is sizable and but is only just beginning to be explored [6, 10, 34].

We also found that our participants reported being surprisingly proactive in engaging with their security and privacy. Indeed, proactive triggers were the second most frequently reported triggers leading to S&P behaviors. Also surprisingly, forced triggers were least frequent. This duality of results is promising, in theory — we want people to be more proactive about S&P and to avoid forcing compliant behaviors. However, since few people use two-factor authentication [33] or password managers [36] or regularly update their software [29], there is clearly much room for improvement.

One limitation in interpreting these results is that because people could select multiple triggers leading up to a behavior change, it's difficult to say which trigger played the most important role in convincing someone to change their behavior. Accordingly, the best we know is that these triggers could have played *some* role. A more general limitation is that because participants may have been several months removed from the event, their memory of the relative order of these triggers and their behavior may be muddled.

### 5.2 RQ2: Individual differences

We next empirically modeled how S&P behavioral triggers varied across individuals and behaviors using a series of random-intercepts logistic regressions. Specifically, we modeled how likely a participant was to report a social trigger, a forced trigger and a proactive trigger given their age, gender, nationality, security behavioral intention and the type of behavior they reported having changed. Due to location restrictions of the AMT platform, we filtered out 83 participants who did not identify as being from the U.S. or India as we did not have enough enough data for other nationalities.

We used a random-intercepts term for each participant to account for repeated observations. We calculated the six pairwise comparisons between the four different behaviors using a contrast matrix with R's multcomp package [27]. Significance levels were corrected using the Bonferroni method. Table 3 shows the results.

Coefficients for the numeric covariates (i.e., age, SBI) indicate a change in log odds that a participant reported a particular trigger leading up a behavior change. A positive coefficient implies that the log odds of a participant reporting a particular trigger increases as the predictor variable increases by one standard deviation, while a negative coefficient implies the opposite. For example, the social trigger regression in Table 3 shows that age has a negative coefficient ($b_{age}^{social} = -0.10$). Thus, for every one-standard deviation increase in age, the model estimates that a participant's log odds to have reported a social trigger should decrease by 0.10 (i.e., younger people are more likely to report social triggers).

For categorical covariates (i.e., behavior types, gender, nationality), coefficients represent the difference in log odds to have experienced a particular trigger between participants at different levels of the covariate. For example, Table 3 shows that the coefficient for a participant from the U.S. to report a proactive trigger versus a participant from India is $b_{US}^{proactive} = 0.81$. As the coefficient is positive and large, we

| Behavior trigger | Abbrev | Type | Mobile Auth | App Uninstall | Password Update | Facebook Privacy |
|---|---|---|---|---|---|---|
| I directly experienced a security breach from someone I know | Breach by Known | Social | 5% | 6% | 4% | 5% |
| I allowed someone to use my device or account previously | Shared Access | Social | 15% | N/A | 8% | 21% |
| I observed people around me doing this | Observed Others | Social | 8% | 14% | 7% | 11% |
| Someone I know advised me to do this | Received Advice | Social | 14% | 16% | 9% | 12% |
| Other Social | Other Social | Social | 2% | $< 1\%$ | $< 1\%$ | $< 1\%$ |
| I directly experienced a security breach from a stranger | Breach by Stranger | Forced | 5% | 5% | 9% | 6% |
| My device or account prompted me to do this | Device Prompt | Forced | 9% | 9% | 22% | 9% |
| My organization required me to do this | Org Prompt | Forced | 6% | 3% | 6% | 3% |
| Other Forced | Other Forced | Forced | $<1\%$ | $< 1\%$ | 3% | 2% |
| I looked through settings / options to do this | Browsed Settings | Proactive | 13% | N/A | 5% | 15% |
| Nothing really happened | No Trigger | Proactive | 7% | 26% | 13% | 5% |
| I read a news article about the security vulnerability or recommending a best practice | Read News | Proactive | 15% | 11% | 13% | 12% |
| Other Proactive | Other Proactive | Proactive | $<1\%$ | $< 1\%$ | $< 1\%$ | 3% |

Table 1: Behavioral triggers, classified into three higher-level types: social, forced, and proactive. Trigger rates for each behavior are provided in the last four columns. The dominant trigger(s) for each behavior is highlighted in green.

| | Mob. Auth | App Del. | Change Pwd | FB Priv. | Over-all |
|---|---|---|---|---|---|
| **Social** | 375 (43%) | 131 (43%) | 273 (29%) | 374 (48%) | 1153 (39%) |
| **Forced** | 179 (21%) | 66 (19%) | 375 (39%) | 153 (20%) | 773 (26%) |
| **Pro-active** | 315 (36%) | 122 (36%) | 312 (33%) | 256 (33%) | 1005 (34%) |

Table 2: Trigger frequency across all four S&P behaviors individually and overall. The dominant trigger type for each behavior is highlighted in green.

can conclude that participants from the U.S. are much more likely than participants from India to report a proactive trigger leading up to a S&P behavior.

Generally, Table 3 shows that there are many significant correlations between behaviors, demographics and how likely one is to report a social, forced or proactive trigger leading up to a S&P behavior. We discuss each key finding, in turn.

**Security Behavioral Intention**: There was a strong correlation between SBI and the triggers participants' reported leading up to their behaviors. Unsurprisingly, people with higher SBI were more likely to report proactive triggers ($b_{sbi}^{proactive} = 0.36, p < 0.001$), while people with lower SBI were more likely to report forced triggers ($b_{sbi}^{forced} = -0.25, p < 0.001$). Perhaps most importantly, we also found that people with lower SBI were more likely to report a social trigger ($b_{sbi}^{social} = 0.12, p < 0.05$).

Figure 4 shows the relationship between SBI and a participants' likelihood of reporting a social, forced or proactive trigger. The likelihood is calculated from our estimated random-intercepts logistic regression model, which also takes into account participants age, gender, nationality and the behavior type. The trend lines are fit using a Gaussian Additive Model, which allows us to model non-linearities in the relationship. We can see a clear trend — controlling for all of the other covariates, people with low-to-medium security behavioral intention (-1.5 to 0.5) are much more likely to report a social trigger leading up to a S&P behavior. From Figure 3, we know that most people ($> 65\%$) fall into this low-to-medium range.

Taken together, we found strong empirical evidence sug-

| | Social | Forced | Proactive |
|---|---|---|---|
| **Intercept** | 0.31 | −1.38∗∗ | −1.59∗∗ |
| *Individual comparisons* | | | |
| **SBI** | −0.12∗ | −0.25∗∗ | 0.36∗∗ |
| **Age** | −0.10∗ | −0.06 | 0.15∗ |
| **Male (vs. Female)** | −0.17 | 0.09 | 0.10 |
| **US (vs. India)** | −0.87∗∗ | 0.09 | 0.81∗∗ |
| *Behavior comparisons* | | | |
| **Pwd (vs. App Uninst.)** | −0.44∗ | 1.05∗∗ | −0.24 |
| **MAuth (vs. App Uninst.)** | 0.13 | 0.12 | 0.12 |
| **FB (vs. App Uninst.)** | 0.37∗ | 0.07 | −0.14 |
| **MAuth (vs. Pwd)** | 0.57∗∗ | −0.93∗∗ | 0.35∗ |
| **FB (vs. Pwd)** | 0.81∗∗ | −0.99∗∗ | 0.10 |
| **FB (vs. MAuth)** | 0.24 | −0.06 | −0.26 |

$p < 0.05$ *, $p < 0.001$ **

Table 3: Logistic regression coefficients comparing how often social, forced and proactive triggers were reported as behavioral triggers for different participants and for different behaviors. Bonferonni correction was used to account for multiple testing. Baseline comparison groups are indicated in parentheses for categorical variables. We used R's multcomp package to compute the six pairwise differences for the four behaviors.

gesting that social triggers are *especially* effective S&P behavioral triggers for the majority of people who have low-to-medium security behavioral intention. Yet, to date, most end-user facing security and privacy systems do not take into account social factors or encourage social interaction. A strong implication for design, then, is to create systems that encourage greater social interaction so that it is easier to reach people with low-to-medium SBI.

**Age and Gender**: There were strong correlations between age and the triggers that reportedly lead up to S&P behaviors. Younger people were more likely to report social triggers ($b_{age}^{social} = -0.10, p < 0.05$) and older people were more likely to report proactive triggers ($b_{age}^{proactive} = 0.15, p < 0.01$). We found no significant correlations between gender and behavioral triggers. Additional research may be needed to determine causality, but our results suggest that some level of age-based personalization may be needed to trigger pro-S&P behaviors.

**Nationality**: We found a strong correlation between self-reported nationality and reported S&P behavioral triggers. People from the U.S. were much more likely to report being individually proactive about their security ($b_{U.S.}^{proactive} = 0.81, p < 0.001$), whereas people from India were much more likely to report social triggers ($b_{U.S.}^{social} = -0.87, p < 0.001$).

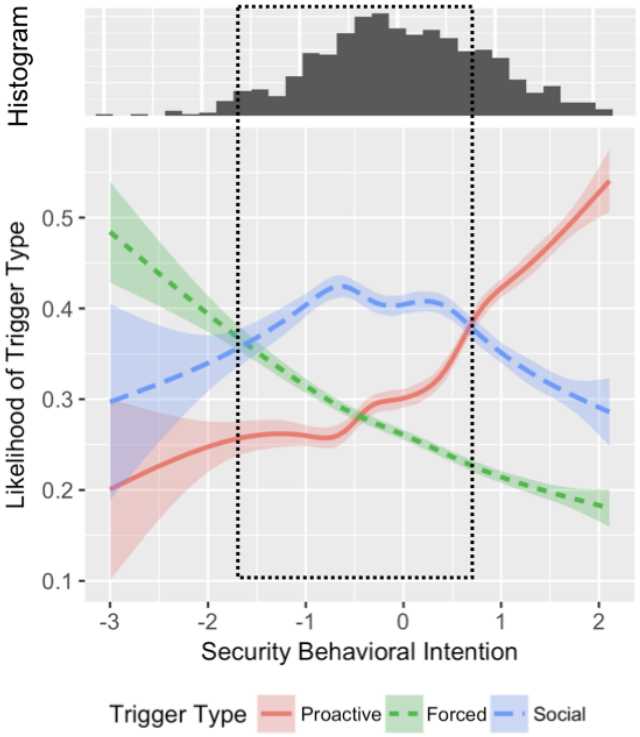



Figure 4: Estimated likelihood of reporting a social, forced or proactive trigger for participants with different SBI. The trend lines, and the 95% confidence intervals, were fit to a Gaussian Additive Model. People with high SBI were more likely to report a proactive trigger, while the majority of people with low-medium SBI were more likely to report social triggers. The dashed boxed outlines the SBI range in which social triggers were most prevalent to facilitate cross-referencing with the SBI histogram above.

Figure 5 visualizes the distribution of the estimated likelihood of reporting a social, forced or proactive trigger for participants from the U.S. vis-a-vis those from India. The estimated likelihoods are calculated from the logistic regression in Table 3, and take into account the other covariates in the regression. We can see a clear separation in the social and proactive distributions, with the former favoring people not from the U.S. and the latter favoring people from the U.S.

These findings echo those of prior work modeling differences in the privacy attitudes of Mechanical Turk workers in India vs. the US [30]. While it's tempting to attribute these effects to cultural differences, our findings do not imply causality. Additional research will be necessary to tease apart the effect of culture from other confounding factors such as, for example, the work contexts of AMT workers in the U.S. versus those in India.

**Behaviors**: Different behaviors had significantly differing trigger distributions. While the raw numbers are pre-

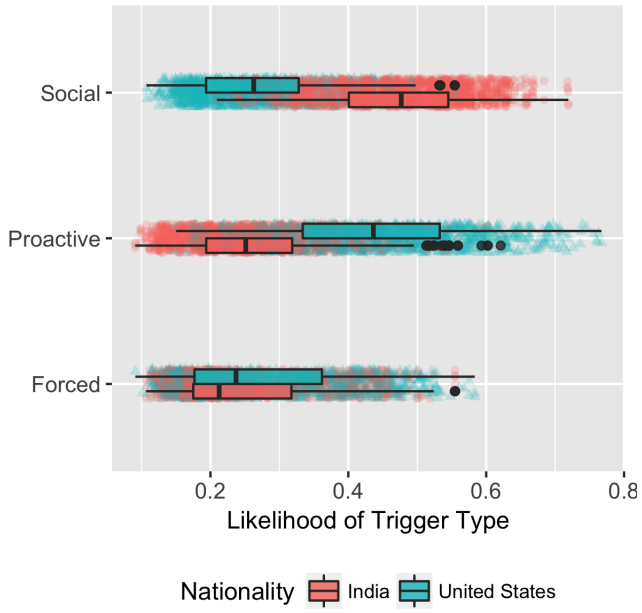



Figure 5: We plot the estimated likelihood that people from the U.S. and from India were to report social, proactive or forced triggers. These likelihoods are estimated from the random-intercepts logistic regressions shown in Table 3, additionally accounting for age, sbi and behavior type. People from the U.S. were more likely to report proactive triggers, while people from India were more like report social triggers.

sented in Table 2, the regression analysis uncovered statistically significant differences across behaviors controlling for age, gender, nationality and SBI. Mobile authentication changes were significantly more likely to have a reported social ($b^{social}_{MobvPwd} = 0.57, p < 0.001$) and proactive ($b^{proactive}_{MobvPwd} = 0.35, p < 0.001$) trigger than changing passwords. Changing passwords was significantly more likely to have a reported forced trigger than mobile authentication changes ($b^{forced}_{MobvPwd} = -0.93, p < 0.001$), changing Facebook privacy settings ($b^{forced}_{FBvPwd} = -0.99, p < 0.001$) and uninstalling applications ($b^{forced}_{PwdvApp} = 1.05, p < 0.001$). Uninstalling applications was more likely to have a reported social trigger than changing passwords ($b^{social}_{PwdvApp} = -0.44, p < 0.01$). Finally, changing Facebook privacy settings was more likely to have a reported social trigger than uninstalling applications ($b^{social}_{FBvApp} = 0.37, p < 0.05$) and changing passwords ($b^{social}_{FBvPwd} = 0.81, p < 0.001$).

## 5.3 RQ3: Sharing patterns

Conversations and interactions about security are rare and avoided by both experts and non-experts alike [9, 11]. Yet, social triggers cannot be produced without some form of active

| | Mobile Auth | App Uninst. | Changed Pwd | FB Priv. | Overall |
|---|---|---|---|---|---|
| **Overall Shared** | 137 (30%) | 173 (46%) | 81 (12%) | 222 (53%) | 613 (32%) |
| **Family** | 66 | 40 | 43 | 64 | 213 |
| **Friend** | 91 | 44 | 82 | 89 | 306 |
| **Colleague** | 42 | 16 | 12 | 12 | 82 |
| **S.O.** | 38 | 17 | 32 | 54 | 141 |
| **Other** | 3 | 1 | 4 | 3 | 11 |

Table 4: Number of people who shared their behavior changes across different behaviors and overall. Participants could select multiple audiences. The first row indicates the total number of those behaviors that were shared.

or passive social interaction. Accordingly, we next wanted to understand when and why people share their security behaviors with others to see if there may be untapped opportunities to encourage greater sharing.

Table 5.3 shows how many participants decided to share their reported behavior changes with others, both overall and with specific other relations (e.g., friends, family, colleagues and significant others). Overall, 32% of reported behavior changes were shared with others — primarily with friends and to a lesser degree with family and significant others. This overall sharing rate is in line with prior work on people's willingness to share news articles about security and privacy, which found that 29% of MTurkers reported sharing such articles with friends and family [12]. We suspect the actual rate of sharing S&P behaviors may be lower in practice, but that the behavior changes participants were reporting on were especially salient and thus more likely to be shared.

We found a large difference in the sharing rate of different S&P behaviors. The most shared behavior was updating one's Facebook privacy settings (53% share rate). This result was unsurprising, given the inherent social nature of Facebook and its salient privacy settings. Conversely, changing passwords was least likely to be shared (12% share rate). This contrast suggests that there remains a significant opportunity to develop systems that encourage more explicit social interactions between individuals, especially for behaviors made outside of a social platform such as Facebook. Indeed, as people with low-to-medium SBI appear to respond especially well to social triggers and are rarely proactive, a high-level goal should be to encourage more social interactions and greater observability of S&P behaviors more generally, albeit with the ability to maintain individual privacy as desired.

We next wanted to explore why people *did* and did *not* share their behaviors with others. If we have a better understanding of the reasons people share their S&P behaviors, we

| | Mobile Auth | App Uninst. | Changed Pwd. | FB Priv. |
|---|---|---|---|---|
| I noticed they were being insecure | 15% | 14% | 12% | 33% |
| They learned about a new security tool | 14% | 9% | 9% | N/A |
| I felt obligated to protect them | 13% | 17% | 18% | N/A |
| They experienced a breach | 12% | 11% | 11% | N/A |
| They had to set up a new device, account or tool | 7% | 4% | 6% | N/A |
| They read a news article about security | 11% | 9% | 8% | 23% |
| I just wanted to talk about my recent change | 15% | 22% | 21% | 43% |
| They noticed that I made a change | 12% | 13% | 12% | N/A |
| No reason | 1% | 0% | 0% | 0% |
| Other | 1% | 2% | 2% | 1% |

Table 5: Reasons people decided to share that they had made a security and privacy behavior change with others. Many people mentioned being vigilant of others' S&P and feeling obligated to protect them. These rows are highlighted in green.

may be able to design targeted systems and interventions that encourage more explicit social interactions. Table 5.3 lists why people elected to share their behaviors with others.

The most commonly reported reason to share was non-descriptive: "I just wanted to talk about my recent change." We included this option for participants who could not select a more specific reason for why they shared their behavior. However, the second and third most commonly reported reasons across all behaviors was that people felt an obligation to protect others and because participants were vigilant of other' being insecure. These findings suggest that people often share their S&P behaviors with others because they feel a sense of accountability or obligation for the security of their friends and loved one, as has been alluded to in past work [9]. However, as has been previously reported, there are very few systems in place that allow people to act on this sense of accountability for their friends and loved ones [11]. Furthermore, the low observability of S&P behaviors places a strong burden on early adopters to explicitly share their behaviors with others if those behaviors are to spread.

| | Coefficent | p-value | |
|---|---|---|---|
| **Intercept** | -0.16 | 0.62 | |
| **Social Trigger?** | 2.31 | <0.001 | ** |
| *Individual differences* | | | |
| **SBI** | 0.07 | 0.41 | |
| **Age** | 0.002 | 0.79 | |
| **Male (vs. Female)** | -0.10 | 0.52 | |
| **US (vs. India)** | -1.10 | <0.001 | ** |
| *Behavior differences* | | | |
| **Pwd (vs. App)** | -2.83 | <0.001 | ** |
| **MAuth (vs. App)** | -1.94 | <0.001 | ** |
| **FB (vs. App)** | -0.65 | 0.01 | * |
| **MAuth (vs. Pwd)** | 0.89 | <0.001 | ** |
| **FB (vs. Pwd)** | 2.19 | <0.001 | ** |
| **FB (vs. Mob)** | 1.23 | <0.001 | ** |

Table 6: Regression coefficients comparing how the decision to share one's new security behavior correlates with one's SBI, demographics, whether or not the behavior was socially triggered, and the type of behavior being shared. Bonferonni correction was applied. Baseline comparison groups are indicated in parentheses for categorical variables. We used R's multcomp package to compute the six pairwise differences for the four behaviors.

The primary reasons *not to share*, unsurprisingly, centered around a general lack of desire to share (38%) and an assumption that other people did not need to know anything about one's S&P behaviors (34%). If we are to increase the prevalence social triggers, these results suggest that we should make S&P systems that encourage social sharing and that are more easily observable so that early adopters do not need to explicitly share their behaviors.

To better understand what factors lead to sharing S&P behaviors, we ran a mixed-effects logistic regression correlating if a participant shared their reported S&P behavior with their age, gender, SBI, nationality, whether or not their behavior was socially triggered, and the type of behavior. The results are shown in Table 6. Coefficients can be interpreted in the same way as in the models reported in Table 3. We found strong, significant correlations as outlined below.

**Nationality**: People from the U.S. were far less likely to share than people from India. ($b^{share}_{U.S.} = -1.10, p < 0.001$). This lack of sharing could also explain the stark difference in social triggers as a catalyst for behavior in the U.S. versus India, but further research is necessary for this to be conclusive.

**Behavior type**: All pairwise differences between the sharing rates of distinct behavior types were significant. Com-

bined with the raw counts of sharing by behavior presented in Table 5.3, it looks like changing Facebook privacy settings is shared most frequently, followed by app uninstallations, mobile authentication changes and, finally, password updates.

**Behavior prompted by a social trigger**: Finally, if participants reported changing their behavior as a result of a social trigger, they were much more likely to share information about that behavior with others ($b^{share}_{social?} = 2.31, p < 0.001$). Concretely, 56% of behaviors that had a reported social trigger were shared with others, compared to just 14% of behaviors that were not — a four-fold increase.

# 6 Discussion

## 6.1 Summary of Results and Contributions

Most generally, we found that social triggers (39%), in which people were influenced by others, were the most frequent reported catalysts for S&P behaviors. Proactive triggers (34%), where people individually decided to make a change independent of an external prompt or breach, were second most frequently reported. Finally, forced triggers (26%), where people made a change in response to a specific breach or news event, were least frequently reported.

While our aggregate results paint a simple picture, once we drilled down into differences between people from different backgrounds and across different behaviors, we uncovered a more nuanced story. Specifically, we found that individual and behavioral differences correlate strongly with which triggers participants reported. Indeed, people with high security behavioral intention were most likely to report proactive triggers, but people with *low-to-medium* SBI, who make up the vast majority, were much more likely to report changing their behavior in response to a social trigger. Demographics also correlated with reported behavioral triggers — younger people and people from India were much more likely to report changing their behavior in response to a social trigger, while older people and people inside the U.S. were much more likely to report changing their behavior proactively.

In analyzing when and why people shared their own security and privacy behaviors with others, we found that people who themselves reported social triggers were far more likely to share their behaviors with others. We also found that people in India were much more likely to share their behaviors with others than people in the U.S., and that different behaviors are shared at different rates — specifically, uninstalling applications for security and privacy reasons was shared most often, followed by updates to Facebook privacy settings, changes to mobile device security and, lastly, password updates.

Finally, we also found that while most people do not share their S&P behaviors with others because they just do not want to, when people *do* share their behaviors they do so because they feel a sense of accountability for or obligation to protect their friends and loved ones.

## 6.2 Design Implications

Our work contributes the first large quantitative analysis comparing the relative frequency of self-reported S&P behavioral triggers and how those triggers vary across individuals from different backgrounds and behavior types. We now reflect on some actionable design implications.

**Designing security and privacy systems that encourage social interaction**: The highest-order bit of our results is a hypothesis — to encourage more widespread use of pro-S&P behaviors by non-experts, these behaviors should be designed to be more passively observable or to encourage greater active social interaction. In other words, we hypothesize that to encourage pro-S&P behaviors, we should design systems and interventions that facilitate social triggers.

The basis for this hypothesis is two-fold: first, social triggers were the most frequently cited prompts for S&P behaviors, in aggregate, and were *especially so* for people with lower security behavioral intention; and, second, people who reported changing their behavior as a result of a social trigger were four times more likely to share their own behaviors with others, in turn. Accordingly, by making more social systems we may be able to bootstrap a feedback loop in which social triggers lead to behavior change, which, in turn, should lead to even more social triggers.

We note that our call to make security more social is not new — prior work has also made similar suggestions [9, 11, 14, 31]. Still, our work adds to an emerging chorus of research illustrating the importance of considering social factors in the design of end-user facing S&P systems.

How can we design systems and interventions that encourage more social sharing? Prior work suggests that by making security systems that are more observable (i.e., a system that is easily seen by others when it is used), cooperative (i.e., a system that allows people to work together towards mutually beneficial ends) and stewarded (i.e., a system that allows one person to act for the benefit of others), people are more likely to both actively engage in social interactions about S&P as well as passively observe others' S&P behaviors [7]. Of course, such systems should also respect the individual privacy preferences of those who would prefer not to be identifiable in social cues to others. For end-user communities who would prefer their individual S&P behaviors to be private, aggregate social cues where no individual is identified may be one effective path forward [10, 14].

**Exploring a broader design space for S&P triggers**: Fogg defines three types of behavioral triggers for persuasive design [20]: *sparks*, which motivate people with high ability but low motivation; *facilitators*, which simplify action for people with high motivation but low ability; and, *signals*, which serve as reminders for people who already have high motivation and ability. Many existing S&P warnings and notifications are *signals*. Sparks and facilitators also pose interesting opportunities for S&P, as few end-users have both

high motivation and high ability to engage in pro-S&P behaviors. An example of a *spark* that encourages S&P behaviors is Das et al.'s social proof notifications, which informed Facebook users of the number of their friends who used optional security tools on Facebook [10]. An example of an effective *facilitator* that simplifies S&P behaviors comes from Akhawe and Felt's redesign of the Chrome SSL warning to simplify exiting out of suspicious webpages [3].

This prior work has only begun to explore a rich design space for sparks and facilitators. We foresee opportunities to co-opt an end-users' social and environmental contexts to create better sparks and facilitators (e.g., trending S&P behaviors in the wake a publicized incident, aggregated social proof cues of others' S&P behaviors in the same room).

**Personalized behavioral triggers:** Our results also illustrate that behavioral triggers may need to be personalized to people from different cultural contexts, demographic backgrounds, and levels of SBI. The growing body of literature on modeling individual differences in S&P paints a nuanced picture of the varying desires, attitudes and assumptions of different groups of people with respect to S&P. Our results inform the need to individually tailor triggers that prompt people to act in a manner consistent with expert S&P advice.

### 6.3 Limitations

As with any study, ours has a number of limitations that are important to keep in mind when interpreting the results. First, our dataset has biases. Specifically, our sample over-represents males and people in technology related fields and occupations. We suspect this is the result of a self-selection bias in who decided to fill out our survey, as other AMT studies tend to have more gender balance (e.g., [12, 30]). The upshot is that our population probably over-represents those with high SBI. In turn, we expect that due to this sample skew, proactive triggers should account for a smaller proportion of reported behavioral triggers in a more representative sample. While these biases are important to consider in interpreting our results, our choice of sample still provides generalizable insights. Indeed, while some prior work suggests that MTurk workers tend to be more concerned about privacy than a U.S.-census representative sample [30], more recent work found that an MTurk sample was more representative of the U.S. population in terms of privacy and security experience, knowledge and advice sources than a census-representative web panel [39].

A second limitation is that our survey is primarily based on self-report and recollection. We asked participants about recent behaviors that occurred but it's likely that their recollection of these behaviors and their triggers is imperfect. This limitation may also contribute to higher-than-expected reporting of proactive triggers — people who cannot recall what factors lead up to a behavior change may simply attribute the change to their own independent judgment. In future work, it would be useful to catch behavior changes closer to the moment those behaviors occur, perhaps through a diary study.

Our data captures a limited subset of S&P behaviors, though we expect it to generalize as well as any other subset. There are many other S&P behaviors we did not ask about — e.g., two-factor authentication enrollment, software updates, and usage of password managers.

The typology of S&P behaviors we explored in this study only capture triggers that are perceived *in-the-moment*. There may be other catalysts for behavior change that play a longer-term role in influencing end-user SP behaviors: e.g., social norms and cultural attitudes.

Finally, our categorization of individual behavioral triggers into "social", "forced" and "proactive" is one of a number of other possible groupings. While our categorization was based on a synthesis of prior work and a thorough discussion amongst the authors of this paper, other groupings of the triggers may also be valid and could, through analysis, offer other insights into S&P behavioral triggers.

## 7 Conclusion

We conducted a large online survey to answer questions about what triggers good S&P behaviors, how that varies across individuals, and how people share their S&P behaviors with others. Social triggers were the most frequently reported behavioral triggers for pro-S&P behaviors, especially among those with low-to-medium security behavioral intention. We also found that participants were four times more likely to share their own S&P behaviors with others when their behaviors were also reported to be socially triggered. This result suggests the possibility of a feedback loop: if we can design behaviors that encourage social interaction, we may be able to trigger additional behavior change which, in turn, should encourage even more social interaction. People from different age groups and nationalities differed in which triggers they reported as prompting their S&P behaviors. Older people and people in the U.S. were more likely to respond to proactive triggers, while younger people and people in India were more likely to respond to social triggers. In summary, we contribute a general typology of in-the-moment, perceived S&P behavioral triggers and identify how those triggers vary across different individuals and behaviors. In turn, this contribution opens up fruitful new opportunities for the design of behavioral triggers meant to encourage pro-S&P behaviors.

## Acknowledgment

This research was generously supported, in part, by the National Science Foundation through grants SaTC-1755625 and CNS-1704087. Tiffany Hyun-Jin Kim made important contributions to this work, particularly in helping design the survey. This work was also significantly improved through a dialogue with our anonymous reviewers, whose effort we appreciate.

## References


[1] Alessandro Acquisti and Jens Grossklags. Losses, Gains, and Hyperbolic Discounting: Privacy Attitudes and Privacy Behavior. In J. Camp and R. Lewis, editors, *The Economics of Information Security*, pages 179–186. 2004.

[2] Anne Adams and Martina Angela Sasse. Users are not the enemy. *Communications of the ACM (CACM)*, 42(12):40–46, dec 1999.

[3] Devdatta Akhawe and Adrienne Porter Felt. Alice in warningland: a large-scale field study of browser security warning effectiveness. In *Proc. USENIX Sec'13*, pages 257–272, 2013.

[4] Monica Anderson and Kenneth Olmstead. Many smartphone users don't take steps to secure their devices. Technical report, Pew Research Center, 2017.

[5] Cristian Bravo-Lillo, Lorrie F. Cranor, Julie Downs, Saranga Komanduri, Robert W. Reeder, Stuart Schechter, and Manya Sleeper. Your Attention Please: Designing security-decision UIs to make genuine risks harder to ignore. In *Proc. SOUPS'13*, 2013.

[6] Lynne M Coventry, Debora Jeske, John M Blythe, James Turland, and Pam Briggs. Personality and social framing in privacy decision-making: A study on cookie acceptance. *Frontiers in psychology*, 7:1341, 2016.

[7] Sauvik Das. Social cybersecurity: Understanding and leveraging social influence to increase security sensitivity. *it - Information Technology*, 58(5):237–245, jan 2016.

[8] Sauvik Das, Eiji Hayashi, and Jason Hong. Exploring Capturable Everyday Memory for Autobiographical Authentication. In *Proc. UbiComp'13*, 2013.

[9] Sauvik Das, Hyun Jin Kim, Laura A. Dabbish, and Jason I. Hong. The Effect of Social Influence on Security Sensitivity. In *Proceedings of the 10th Symposium on Usable Privacy and Security (SOUPS'14)*, 2014.

[10] Sauvik Das, Adam D.I. Kramer, Laura A. Dabbish, and Jason I. Hong. Increasing Security Sensitivity With Social Proof: A Large-Scale Experimental Confirmation. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security (CCS '14)*, pages 739–749, New York, New York, USA, 2014. ACM Press.

[11] Sauvik Das, Adam D.I. Kramer, Laura A. Dabbish, and Jason I. Hong. The Role of Social Influence in Security Feature Adoption. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*, pages 1416–1426, New York, New York, USA, 2015. ACM Press.

[12] Sauvik Das, Joanne Lo, Laura Dabbish, and Jason I Hong. Breaking! A Typology of Security and Privacy News and How It's Shared. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, pages 1–12, New York, New York, USA, 2018. ACM Press.

[13] Rachna Dhamija, J D Tygar, and Marti Hearst. Why phishing works. In *Proc. CHI '06*, number April, pages 581–590, New York, New York, USA, 2006. ACM Press.

[14] Paul DiGioia and Paul Dourish. Social navigation as a model for usable security. In *Proc. SOUPS '05*, pages 101–108, New York, New York, USA, 2005. ACM Press.

[15] Paul Dourish, Rebecca E. Grinter, Jessica Delgado de la Flor, and Melissa Joseph. Security in the wild: user strategies for managing security as an everyday, practical problem. *Personal and Ubiquitous Computing*, 8(6):391–401, sep 2004.

[16] Serge Egelman, Lorrie Faith Cranor, and Jason Hong. You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In *Proc. CHI '08*, page 1065, New York, New York, USA, 2008. ACM Press.

[17] Serge Egelman, David Molnar, Nicolas Christin, Alessandro Acquisti, Cormac Herley, and Shriram Krishnamurthi. Please Continue to Hold: An empirical study on user tolerance of security delays. In *Proc. WEIS'10*, 2010.

[18] Serge Egelman and Eyal Peer. Scaling the Security Wall: Developing a Security Behavior Intentions Scale (SeBIS). In *Proc. CHI'15*, pages 2873–2882, New York, New York, USA, 2015. ACM Press.

[19] Adrienne Porter Felt, Alex Ainslie, Robert W Reeder, Sunny Consolvo, Somas Thyagaraja, Alan Bettes, Helen Harris, and Jeff Grimes. Improving SSL Warnings. In *Proc. CHI'15*, pages 2893–2902, 2015.

[20] BJ Fogg. A behavior model for persuasive design. In *Proceedings of the 4th International Conference on Persuasive Technology - Persuasive '09*, page 1.

[21] SM Furnell, A Jusoh, and D Katsabas. The challenges of undersatnding and using security: A survey of end-users. *Computers & Security*, 25(1):27–35, 2006.

[22] Shirley Gaw, Edward W Felten, and Patricia Fernandez-Kelly. Secrecy, Flagging, and Paranoia: Adoption Criteria in Encrypted E-Mail. In *Proceedings of the SIGCHI*

*conference on Human Factors in computing systems (CHI '06)*, pages 591–600, New York, New York, USA, 2006. ACM Press.

[23] Joseph K. Goodman, Cynthia E. Cryder, and Amar Cheema. Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples. *Journal of Behavioral Decision Making*, 26(3):213–224, 2013.

[24] Marian Harbach, Emanuel von Zezschwitz, Andreas Fichtner, Alexander De Luca, and Matthew Smith. It's a Hard Lock Life: A Field Study of Smartphone (Un)Locking Behavior and Risk Perception. In *SOUPS '14: Proceedings of the Tenth Symposium On Usable Privacy and Security*, pages 213–230, 2016.

[25] Cormac Herley. So long, and no thanks for the externalities. In *Proc. NSPW '09*, pages 133–144, New York, New York, USA, 2009. ACM Press.

[26] Cormac Herley. Unfalsifiability of security claims. *Proceedings of the National Academy of Sciences*, 113(23):6415–6420, 2016.

[27] Torsten Hothorn, Frank Bretz, Peter Westfall, Richard M. Heiberger, Andre Schuetzenmeister, and Susan Scheibe. Simultaneous Inference in General Parametric Models, 2017.

[28] Philip G. Inglesant and M Angela Sasse. The true cost of unusable password policies. In *Proc. CHI'10*, pages 383–392, New York, New York, USA, 2010. ACM Press.

[29] Iulia Ion, Rob Reeder, and Sunny Consolvo. "...no one can hack my mind": Comparing Expert and Non-Expert Security Practices. In *Symposium on Usable Privacy and Security (SOUPS)*, pages 327–346, jan 2015.

[30] Ruogu Kang, Stephanie Brown, Laura Dabbish, and Sara Kiesler. Privacy Attitudes of Mechanical Turk Workers and the U.S. Public. In *Proc. SOUPS'14*, pages 37–49, 2014.

[31] Heather Richter Lipford and Mary Ellen Zurko. Someone to watch over me. In *Proceedings of the 2012 workshop on New security paradigms - NSPW '12*, page 67, 2012.

[32] Alexander De Luca, Sauvik Das, Martin Ortlieb, Iulia Ion, and Ben Laurie. Expert and Non-Expert Attitudes towards ( Secure ) Instant Messaging. In *Proceedings of the Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*, 2016.

[33] Grzegorz Milka. The Anatomy of Account Take-Over. In *USENIX ENIGMA*, 2018.

[34] James Nicholson, Lynne Coventry, and Pam Briggs. Can we fight social engineering attacks by social means? assessing social salience as a means to improve phish detection. In *Thirteenth Symposium on Usable Privacy and Security ({SOUPS} 2017)*, pages 285–298, 2017.

[35] Donald A Norman and Stephen W Draper. *User centered system design: New perspectives on human-computer interaction*. CRC Press, 1986.

[36] Kenneth Olmstead and Aaron Smith. Americans and Cybersecurity. Technical report, Pew Research Center, 2017.

[37] Emilee Rader, Rick Wash, and Brandon Brooks. Stories as informal lessons about security. In *Proc. SOUPS '12*, New York, New York, USA, 2012. ACM Press.

[38] Elissa M Redmiles, Sean Kross, and Michelle L Mazurek. How I Learned to be Secure: A Census-Representative Survey of Security Advice Sources and Behavior. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security - CCS'16*, pages 666–677, New York, New York, USA, 2016. ACM Press.

[39] Elissa M Redmiles, Sean Kross, and Michelle L Mazurek. How well do my results generalize? comparing security and privacy survey results from mturk, web, and telephone samples. In *IEEE Symposium on Security Privacy (S&P'19)*, page 0. IEEE, 2019.

[40] Elissa M. Redmiles, Amelia R. Malone, and Michelle L. Mazurek. I Think They're Trying to Tell Me Something: Advice Sources and Selection for Digital Security. In *2016 IEEE Symposium on Security and Privacy (SP)*, pages 272–288. IEEE, may 2016.

[41] M.A. Sasse. Computer security: Anatomy of a Usability Disaster, and a Plan for Recovery. In *Proc. CHI Workshop on HCI and Security Systems*. Citeseer, 2003.

[42] Florian Schaub, Rebecca Balebako, Adam L Durity, and Lorrie Faith Cranor. A design space for effective privacy notices. In *Eleventh Symposium On Usable Privacy and Security ({SOUPS} 2015)*, pages 1–17, 2015.

[43] Manya Sleeper, Rebecca Balebako, Sauvik Das, Amber Lynn McConahy, Jason Wiese, and Lorrie Faith Cranor. The Post that Wasn't: Exploring Self-Censorship on Facebook. In *Proceedings of the 2013 conference on Computer supported cooperative work - CSCW '13*, pages 793–802, New York, New York, USA, 2013. ACM Press.

[44] JM Stanton, P Mastrangelo, KR Stam, and Jeffrey Jolton. Behavioral Information Security: Two End User Survey Studies of Motivation and Security Practices. *AMCIS*, (August):2–8, 2004.

[45] Yang Wang, Gregory Norcie, Saranga Komanduri, Alessandro Acquisti, Pedro Giovanni Leon, and Lorrie Faith Cranor. "I regretted the minute I pressed share": A Qualitative STudy of Regrets on Facebook. In *Proc. SOUPS 2011*, page 1, New York, New York, USA, 2011. ACM Press.

[46] Rick Wash. Folk models of home computer security. In *Proc. SOUPS '10*, page 1, New York, New York, USA, 2010. ACM Press.

[47] Alma Whitten and J.D. Tygar. Why Johnny can't encrypt: A usability evaluation of PGP 5.0. In *Proc. SSYM'99*, pages 14–28, 1999.

## A Survey Questionnaire

**Page 1:** Shown to all participants.

1. What type of a cell phone do you have?
    - iPhone, Android, or other Smartphone
    - Non-smartphone cell phone
    - I don't know what kind of a cell phone I have
    - I don't own a cell phone

**Page 2**: Shown only to participants who selected 'iPhone, Android or other smartphone' in Page 1, Question 1 (P1Q1).

1. How do you mainly use your phone? Select all that apply.
    - Make phone calls
    - Check emails
    - Access social networking sites, such as Facebook, Twitter, Instagram, etc.
    - Access the Internet
    - Shopping, such as Amazon, Netflix, iTunes, etc.
    - Banking
    - Play games
    - Other: *Manual write-in*
2. Have you done any of the following in the past 6 months? Check all that apply.
    - Enabled or changed authentication on any of your mobile devices (e.g., 4-digit PIN, Android 9-dot, password, fingerprint, face recognition on your phone, laptop, tablet or other portable electronic device)
    - Updated your Facebook privacy settings
    - Uninstalled a smartphone app for privacy or security reasons
    - Changed a password on an online account

**Page 3:** Shown to all participants

*This study requires you to share your opinion. It is important that you take the time to read all instructions and questions carefully before you answer them. Previous research has found that some people do not take time to read everything that is displayed in the questionnaire. The questions below serve to test whether you actually take time to do so. If you read this, please answer 'two' on the question 4, add two to that number and use the result as the answer on question 5. Thank you for participating and taking time to read all instructions.*

1. How many email addresses do you maintain?
2. How many social media accounts do you maintain?

**Page 4:** Shown to participants who fail P3 attention checks.

1. Is there someone to whom you truly want to talk about the recent change? Previous research has found that some people do not take time to read questions and answer options carefully. This question serves to test whether you actually take the time to do so. If you read this, please select 'colleague'. Thank you for taking time to read all instructions.
    - Friend
    - Family member
    - Significant other
    - Colleague
    - Other
    - None of the above

*If this attention check is also failed, participants are disqualified and sent to the final page.*

**Page 5:** Shown to participants who did not recall engaging in any of the behaviors listed in P2Q2.

1. Do you recall the most recent security or privacy behavior that you have changed on your mobile device or on the Internet? Please describe it briefly.
    - *Open response*

**Page 6:** Shown for each behavior participants selected in P2Q2 or P5Q1.

1. (Brief description of behavior being asked about). Did any of the following happen before you made the change? Please select all that apply.

- I directly experienced a security breach from a stranger
- I directly experienced a security breach from someone I know
- I allowed someone to use my device / account previously
- *(Facebook privacy update only)* I noticed that my Facebook activities were visible to unintended people
- *(App uninstallation only)* I noticed that the app required unusual permissions
- I observed / heard about other people doing this
- Someone I know advised me to do this
- The device prompted me to do this prior to use
- My organization required me to do this
- I read a news article about the security vulnerability or recommending a best practice
- I looked through settings/options for my mobile device
- Other (required): *Manual write-in*
- Nothing in particular happened before this change

2. (if selected option: 'I directly experienced a security breach from someone I know' in P6Q1) Who breached security on you? Please select all that apply.
   - Friend
   - Family member
   - Significant other (spouse / boyfriend / girlfriend)
   - Colleague
   - Other (required): *Manual write-in*
   - I don't remember
3. (if selected option: 'I allowed someone to use my device / account previously' in P6Q1) Who used your device / account previously? Please select all that apply.
   - Friend
   - Family member
   - Significant other (spouse / boyfriend / girlfriend)
   - Colleague
   - Other (required): *Manual write-in*
   - I don't remember
4. (if selected option: 'I allowed someone to use my device / account previously' in P6Q1) Who used your device / account previously? Please select all that apply.
   - Friend
   - Family member
   - Significant other (spouse / boyfriend / girlfriend)
   - Colleague
   - Other (required): *Manual write-in*
   - I don't remember
5. (if selected option: 'I observed people around me doing this' in P6Q1). You observed people around you doing this. Who did you observe? Please select all that apply.
   - Friend
   - Family member
   - Significant other (spouse / boyfriend / girlfriend)
   - Colleague
   - Other (required): *Manual write-in*
   - I don't remember
6. (if selected option: 'Someone I know advised me to do this' in P6Q1). Who advised you to make this change? Please select all that apply.
   - Friend
   - Family member
   - Significant other (spouse / boyfriend / girlfriend)
   - Colleague
   - Other (required): *Manual write-in*
   - I don't remember

**Page 7:** Shown for each behavior participants selected in P2Q2 or P5Q1.

1. After you made the change, did you talk about it to anyone else? Who did you talk with most recently?
   - Friend
   - Family member
   - Significant other (spouse / boyfriend / girlfriend)
   - Colleague
   - Other (required): *Manual write-in*
   - I didn't talk about this with anyone.
2. (if selected option 'I didn't talk about this with anyone' in P7Q1) Why did you decide not to talk about this to anyone? Please select all that apply.
   - I didn't feel comfortable to talk about security
   - I assumed that people already knew about this
   - I assumed that people didn't need to know about this
   - I just didn't want to talk about this to anyone

- I didn't have a chance to talk about this to anyone yet
- Other (required): *Manual write-in*

3. (if selection any option *except* 'I didn't talk about this with anyone' in P7Q1) What channel did you use to talk about the change most recently?
   - Face to face conversation
   - Phone call
   - Text message or email
   - Facebook
   - Twitter
   - Other (required): *Manual write-in*
4. (if selection any option *except* 'I didn't talk about this with anyone' in P7Q1) What prompted you to talk about the change with them? Please select all that apply.
   - I noticed they were being insecure
   - They noticed my change
   - They learned about a new security tool
   - I felt obligated to protect them
   - They experienced a security or privacy breach
   - They had to set up a new device, account, or security tool
   - They read a news article about security
   - I just wanted to talk about my recent change
   - Other: *Manual write-in*
   - None of the above
5. (if selection any option *except* 'I didn't talk about this with anyone' in P7Q1) What did you talk about in your conversation? Please select all that apply.
   - I shared a notification or warning of a potential security or privacy threat
   - I demonstrated insecure behavior
   - I shared instructions on how to change insecure behavior
   - I shared specific advice
   - I shared a story about an experience I had
   - I shared my emotional venting
   - I just talked about a security event
   - Other: *Manual write-in*
   - I just talked about the change I made

**Page 8:** Shown to all participants who passed the attention checks.

1. How would you evaluate your computer literacy level?
   - Very low: I don't know much about computers (1)
   - Low (2)
   - Neither high nor low (3)
   - High (4)
   - Very high: I know a lot about computers (5)
2. How would you evaluate your Internet literacy level?
   - Very low: I don't know much about how the Internet works (1)
   - Low (2)
   - Neither high nor low (3)
   - High (4)
   - Very high: I know a lot about how the Internet works (5)
3. How many hours per week are you on the Internet for reasons other than work (both using the smartphone, tablets, or computers)?
   - 0 to 10 hours
   - 10 to 20 hours
   - 20 to 30 hours
   - 30 to 40 hours
   - More than 40 hours
4. How many different online communities (e.g., reddit), social networks (e.g., Facebook), or online groups (e.g., email list) do you read or post in regularly?
   - None
   - 1
   - 2 to 4
   - 5 or more
5. How many hours per day do you spend on sharing and reading content on social networking sites (e.g., Facebook, Twitter, Google+, Instagram, etc.)?
   - 0 to 1 hour
   - 1 to 3 hours
   - 3 to 6 hours
   - 6 to 9 hours
   - More than 9 hours
6. Please rate your familiarity with the following concepts or tools on the following scale:
   - I never heard about this
   - I heard about this but I don't know what it is

- I know what this is but I don't know how it works
- I know generally how it works
- I know very well how this works

(a) IP address

(b) Cookie

(c) Secure Socket Layer (SSL) / Transport Layer Security (TLS)

(d) Virtual Private Network (VPN)

(e) Encryption

(f) Proxy server

(g) Tor

(h) Privacy settings for your web browser

(i) Private browsing mode in browsers

7. Please indicate whether you think each statement is true or false. Please select "I'm not sure" if you don't know the answer.

   - Private browsing mode in browsers prevents websites from collecting information about you.
   - Login cookies can store username/id and a random string in your web browser to keep the user signed in.
   - No one, except for the sender and intended receiver, can reveal the content of an encrypted message.
   - Tor can be used to hide the source of a network request from the destination.
   - A Virtual Private Network (VPN) is the same as a Proxy server.
   - IP addresses can always uniquely identify your computer.
   - HTTPS is standard HTTP with SSL / TLS to preserve the confidentiality of network traffic.
   - A proxy server cannot be tracked to the original source.

**Page 9:** Shown to all participants who passed the attention checks. SEBIS questions from Egelman and Peer [18].

1. Please indicate how often you have done the following on the following scale:

   - Never
   - Rarely
   - Sometimes
   - Often
   - Always

(a) I set my computer screen to automatically lock if I don't use it for a prolonged period of time.

(b) I use a password/passcode to unlock my laptop or tablet.

(c) I manually lock my computer screen when I step away from it.

(d) I use a PIN or passcode to unlock my mobile phone.

(e) I do not change my passwords, unless I have to.

(f) I use different passwords for different accounts that I have.

(g) When I create a new online account, I try to use a password that goes beyond the site's minimum requirements.

(h) I do not include special characters in my password if it's not required.

(i) When someone sends me a link, I open it without first verifying where it goes.

(j) I know what website I'm visiting based on its look and feel, rather than by looking at the URL bar.

(k) I submit information to websites without first verifying that it will be sent securely (e.g., SSL, "https://", a lock icon).

(l) When browsing websites, I mouseover links to see where they go, before clicking them.

(m) If I discover a security problem, I continue what I was doing because I assume someone else will fix it.

(n) When I'm prompted about a software update, I install it right away.

(o) I try to make sure that the programs I use are up-to-date.

(p) I verify that my anti-virus software has been regularly updating itself.

**Page 10:** Shown to all participants who passed the attention checks.

1. While using the Internet, have you ever done any of the following? Please check all that apply.

   - I have used a temporary username or email address.
   - I have used a fake name or username.
   - I have given inaccurate or misleading information about myself.
   - I have set my browser to disable or turn off cookies.
   - I have cleared cookies and browser history.

- I have used a service that helped me browse the web anonymously, such as a proxy server, Tor, or a virtual personal network (VPN).
- I have sent encrypted e-mails.
- I have decided not to use a website because they asked for my real name.
- I have deleted something I posted in the past.
- I have asked someone to remove something that was posted about me online.
- I have used a public computer to browse anonymously.

2. If we ask you to perform the following actions now, can you do it without getting help from others? Please answer on the following scale.
   - Yes I can do this without getting help from others
   - Probably but I may need help from time to time
   - No I need help from others to do this

   (a) Change authentication on mobile devices
   (b) Change Facebook privacy settings
   (c) Change passwords of your online account
   (d) Check permission requests when downloading an app on mobile devices

3. Have you ever done any of the following? Please select all that apply.
   - I have turned off the automatic connections to free Wi-Fi on my mobile device(s)
   - I have looked for "https" when browsing or shopping on my mobile devices
   - I have turned on login approvals on my Facebook account
   - I have enabled secure browsing on my Facebook account
   - I have kept the same password for an online account after logging in using a public computer
   - I have clicked a URL link on an email and entered my username and password

**Page 11:** Shown to all participants who passed the attention checks.

1. What is your gender?*
   - Male
   - Female
   - Non-conforming
   - Prefer not to answer
2. What is your age?
3. What is your current relationship status?
4. Are you a parent or guardian of any children under 18 years of age?
5. How many adults (age 18 or older) currently live in your household, including yourself? *Optional manual write-in*
6. What is the highest level of school you have completed or the highest degree you have received?
7. Out of the following, which best describes your major (if you are a student) or occupation (if you are a professional)?*
   - Cybersecurity related
   - Computer Science related
   - Other engineering or technology related
   - Other: *Manual write-in*
8. What nationality do you most identify with?
9. Do you have native or fluent proficiency in English?
10. Are you Hispanic or Latino?
11. What is your race? Please select all that apply.
    - American Indian or Alaska Native
    - Asian
    - Black or African American
    - Native Hawaiian or other Pacific Islander
    - White
    - Prefer not to answer

# Replication: No One Can Hack My Mind
# Revisiting a Study on Expert and Non-Expert Security Practices and Advice

Karoline Busse
*University of Bonn*
*busse@cs.uni-bonn.de*

Julia Schäfer
*Univeristy of Bonn*
*s6juscha@gmail.com*

Matthew Smith
*University of Bonn / Fraunhofer FKIE*
*smith@cs.uni-bonn.de*

## Abstract

A 2015 study by Iulia Ion, Rob Reeder, and Sunny Consolvo examined the self-reported security behavior of security experts and non-experts. They also analyzed what kind of security advice experts gave to non-experts and how realistic and effective they think typical advice is.

Now, roughly four years later, we aimed to replicate and extend this study with a similar set of non-experts and a different set of experts. For the non-experts, we recruited 288 MTurk participants, just as Ion et al. did. We also recruited 75 mostly European security experts, in contrast to the mostly US sample from Ion et al. Our findings show that despite the different samples and the four years that have passed, the most common pieces of expert advice are mostly unchanged, with one notable exception. In addition, we did see a fair amount of fluctuation in the long tail of advice. Non-expert self-reported behavior, however, is unchanged, meaning that the gap between experts and non-experts seen in Ion et al.'s work is still just as prominent in our study. To extend the work, we also conducted an A/B study to get a better understanding of one of the key questions concerning experts' recommendations, and we identified types of advice where research by the usable security community is most sorely needed.

## 1 Introduction

Whenever the media picks up on the latest data breach, various sources seize the opportunity to give advice such as "Do not use the same passwords for all systems" [9] or "Antivirus software is crucial to protecting your computer." [23] Under this barrage of different advice, selecting and following "good" advice is a difficult task for users [10]. Factors such as socioeconomic status, consumer habits, or conveniences also play a role in the decision-making process [24, 26]. Even when advice is regarded as "good" by a user, it is not necessarily a given that they know how to apply it in their own individual context. We must not overlook the limits of users' capability taking into account the complexity of any advice we give [5].

In 2015, Ion, Reeder, and Consolvo explored the opinions and beliefs of expert and non-expert users in a survey study and found that users neglect three vital security practices that experts strongly advise: installing software updates, using two-factor authentication, and using a password manager. On the other side, non-experts regarded antivirus software as a very important security practice, unlike the experts, who were not convinced by it. Almost four years have passed since that study, which is a long period of time in terms of technological innovation and security practices. Security and privacy continue to gain more widespread recognition, so we were interested to see what, if anything, had changed with respect to expert advice and non-expert self-reported behavior.

We thus conducted two online surveys, one for experts and one for non-experts, and compared the results to the previous study by Ion et al. Many of the past security topics and advice covered in the original work are still relevant today. We also discovered that some of the topics relevant to users in the past have been replaced by newer topics, for example, the spread of blocking extensions for web browsers, which are able to manage cookies. Where in the past, users were concerned with regularly deleting cookies, they now rely on blocking extensions.

Apart from seeing if and how our sample differed from the original, we wanted to explore a methodological issue in the original study. One of the central parts of the original study concerned how effective *and* realistic particular types of advice are. This information from experts was gathered using compound questions, and the advice was ranked and compared on that basis. Compound questions can be problematic because it is not clear how participants combine the separate


*USENIX Symposium on Usable Privacy and Security (SOUPS)* 2019.
August 11–13, 2019, Santa Clara, CA, USA.

components [4]. For example, when asked to rank advice on a five-point scale, a 3 could mean an expert thought that a piece of advice was extremely effective (5) but completely unrealistic (1), or vice versa, and the expert combined the two values into a simple average. However, a 3 could also be given because the expert thought the piece of advice was a 3 regarding realism and a 3 in effectiveness. To make matters worse, the same separate assessment from above (extremely effective (5) but completely unrealistic (1)) could also be combined by the expert into a 1 if the expert takes the view that if a piece of advice is unrealistic, then the combined effectiveness is also a 1 . So the same separate assessments can lead to very different combined scores and separate results from the same assessments can lead to very different combined scores.

While the combined score is useful because it reflects the personal assessment of an expert participant using whatever weighted combination they deem most appropriate, it potentially hides interesting discrepancies that could highlight which pieces of advice could be particularly important for researchers to improve and, more specifically, which areas need improvement. For example, a piece of advice that gets a 5 for effectiveness but a 1 for realism is probably a good candidate for researchers to improve the usability. On the other hand, a 4 on realism and a 2 on effectiveness could indicate that systems research is needed to improve effectiveness or it might be best if the advice is discouraged, since it uses up valuable security budget without being particularly effective. To be able to compare our data directly with the original work by Ion et al., in addition to gaining the insights described above, we gave half our expert participants the original compound questions and half the experts got the questions broken down into their compound elements.

Based on our analysis, we suggest four fields where usable security research is needed to improve existing methods or invent new ways of handling the implied security issues. The areas are: password security, two-factor authentication, links and attachments, as well as application updates. Out of these four fields, three were already prominently discussed in the original work, suggesting that the research and engineering communities in usable security still have a lot of work to do.

The remainder of this paper is structured as follows: Section 2 gives an overview of relevant work regarding security and privacy advice, as well as an in-depth look at the original study by Ion et al. Section 3 documents our survey methodology for both the expert and non-expert surveys and discusses the design changes we made. In Section 4, we present our replication results and compare them to the original work. The discussion of results, replication efforts, design changes, and fields of action follows in Section 5. We conclude by outlining the limitations of our study (Section 6) and summarizing our work's contributions in Section 7.

## 2 Related Work

In 2008, MacGeorge et al. proposed that for recipients to follow good advice, it should: be useful, comprehensible, and relevant; be effective at addressing the problem; be likely to be accomplished by the recipient; and not possess too many limitations and drawbacks. When giving advice, experts should make sure that the advice is solicited by the recipient, they only give advice if they are a qualified source on the topic; they consider the recipient's point of view; and they exercise sensitivity in phrasing and formulation [20].

Redmiles et al. researched which kinds of advice users adopted and which they rejected. They found that IT professionals, the workplace environment, and negative events, whether personally experienced or told by news media, are users' main sources of digital security advice [26]. As a result of being unable to evaluate the content of a piece of advice, users tend to wager the acceptance of advice based on the trustworthiness of the source. Rejection of advice is influenced by many factors, such as believing that the responsibility for security lies with someone else, perceiving that the advice contains too much marketing material, or believing that the advice might threaten the user's privacy.

In a follow-up US-representative survey on security advice and trusted sources in 2016, Redmiles et al. identified a digital security divide along lines of the socioeconomic status of participants. Wealthier people tended to have better skills and acquired advice from the workplace, while disadvantaged users relied on family and friends for advice [24].

A Pew Research study by Lenhart et al. investigated where teens between the ages of 12 and 17 get their privacy advice from [17]. A focus group study revealed that teens mainly research and iterate through privacy settings on their own, while a follow-up survey suggests that they also relied on personal advice from friends, parents, or siblings. In general, younger teens relied more on interpersonal advice, while older teens tried to figure things out for themselves.

Harbach et al. explicated in a 2014 survey that risk awareness is often the primary stage for the adoption of security mechanisms and their interactions [13]. While being an essential part of the study of human aspects of security research, it needs to be explored in detail in the context of users' daily lives. A fundamental part of devising usable IT security mechanisms is evaluating which risks and consequences are known to users and, therefore, are already accounted for in their mental budget of coping with security behaviors.

Wash researched so-called *folk models* of home computer users, conducting a series of interviews to identify common models about security threats, namely hackers and viruses. After identifying four virus and four hacker models, Wash set them in relation to popular security advice and suggested which type of user would react in what fashion to each individual piece of advice. This gives a possible explanation for why users do not follow security advice given by experts [34].

Fagan and Khan further investigated why some users follow advice and others do not. They conducted a survey study where they asked participants about their motivations regarding (not) updating, using a password manager, using two-factor authentication, and changing passwords frequently. The authors determined that following security advice was mainly a trade-off decision between convenience and security, where users actively considered features such as set-up time and weighed that against the potential security benefits [10].

## 2.1 The Original Study

In 2015, Iulia Ion, Rob Reeder, and Sunny Consolvo presented their survey-based study on the differences and similarities in online security-related behavior of expert and non-expert users [15]. They developed a four-part survey asking about top security advice and the respondent's own security and privacy habits, as well as asking respondents to rate pre-formulated advice statements for their effectiveness and practicability.

The two surveys that make up the core of their study are based on data gathered by conducting semi-structured interviews with 40 security experts at the 2013 BlackHat, DefCon, and USENIX security conferences.

The expert survey, crafted from the information gathered in the preliminary interviews at security conferences, was conducted from February to April 2014. A minimum of 5 years of work experience in a security-related field was required to be counted as an "expert." Participants were recruited through a post on the Google Online Security Blog [28] and social media. The survey first asked participants to enter three pieces of advice for non tech-savvy users and the three things the participants do themselves to protect their security online. The second part consisted of multiple-choice questions inquiring on certain security-related behaviors and practices. The main part asked the participants to rate pieces of advice directed at non-tech-savvy users. Experts were then asked to rate each piece of advice with regard to both the advice's effect on security and the probability that the user would follow the advice. The survey closed with demographic questions. 231 participants met the criteria for being an expert of working or studying in a security-related field for at least five years.

The non-expert survey was conducted with 294 US-based participants recruited via Amazon Mechanical Turk (MTurk).

The results showed that experts and non-experts followed different approaches to protecting their security online, with the practice of using strong passwords being the only commonality for both groups, ranking in the top 5 responses to the question about the respondents' personal top three security practices (cf. Figure 1). The security practices mentioned by experts were consistent with the experts' ratings of different pieces of advice. These pieces of advice were grouped into four categories: *software updates*, *antivirus software*, *password management*, and *mindfulness*. The security practices utilized by the non-experts received mixed ratings from the experts. Some non-expert practices were considered by the experts to be a good practice, like installing antivirus software and using strong passwords. However, the non-experts' failure to comply with some practices were considered bad habits by the experts, including failure to delete cookies and failure to visit only known websites, among others.

The authors found three security practices that experts followed and recommended that were not employed by the non-experts (see Figure 3), namely installing system updates, using a password manager, and using two-factor authentication, which were considered most important by a majority of the experts. Their results suggest that a combination of better communication and improvements in the systems and their usability were necessary to get non-experts to adhere to these three security practices.

# 3 Methodology

The authors of the original study shared their study materials with us so that we could recreate the surveys as precisely as possible. They also shared the data shown in Figure 1 from their original paper; however, the raw data could not be shared.

The questionnaire featured mostly closed questions that allowed participants to enter free-text data in an "other" answer option. The questions on the practicability of advice with featured the compound design in the original study were 5 point Likert-scale item batteries with optional free text comment fields in between. Our split-question design thus increased the number of questions for participants who answered our modified survey.

The full questionnaires can be found in the appendix A. In total we had three different questionnaires: the expert and end-user questionnaires from the original study and our modified expert questionnaire which separated the compound questions. All questionnaires as well as the pre-study interviews started by getting informed consent. Audio recordings were made in the pre-study with participant consent and then stored on encrypted storage and deleted after evaluation. In compliance with the EU-GDPR, we did not store any personal identifying data such as IP addresses for any online survey.

The responses to the open-ended questions regarding the top three pieces of security advice and the top three personal security practices of experts were coded by two of the authors. First, both researchers coded the results independently and then codes were compared and differences were discussed. Since the coding was straight-forward, full agreement on the codes was reached.

## 3.1 End User Survey

We replicated the end user survey with the same MTurk recruitment criteria as the original authors used: Participants

were required to be from the United States, have a task approval rate of 95% or better and have completed at least 500 tasks. For the sake of replication, we advertised the study with the original payment of 1$, but for fairness reasons we awarded an additional 2$ through MTurk's worker bonus system after the study was concluded. The study was conducted in May 2018.

### 3.2 Expert Interviews and Survey

Based on the expert survey from Ion et al., we conducted 40 interviews with IT security experts at the CeBIT international trade fair on information technology in 2018. Our goal was to evaluate the survey design and gather first impressions for the experts group.

During the course of the interviews, it became clear that the compound question regarding the evaluation of advice[1] led to confusion and insecurities in participants. They often misinterpreted or morphed the question's phrasing after rating a couple of items, leading to decreased comparability of results.

We discussed this finding and the problem of compound questions with the authors of the original study. They chose the compound question due to time constraints. Their pretesting suggested that the length of the survey had to be limited and thus this compromise was made. Also, they were mainly interested in what the experts' overall assessment of advice was and thus the separate components were not as relevant for their work.

Nonetheless, compound questions can be tricky to interpret and important nuances can be lost. In particular, we thought it would be valuable to see if there are any pieces of advice where effectiveness and realism diverge, since these could highlight areas of improvement.

To this end, we separated the compound rating tasks for advice effectiveness and realism. Since this is a divergence from the replication, we assigned half the participants to this survey and the other half completed the original survey with the compound questions. We chose a between-groups design over a within-groups one because we wanted to limit fatigue effects, as the survey was already rather long and repetitive. In addition, we randomized the order of appearance of individual advice items within the 5-piece rating blocks for both groups (see Appendix A) to minimize cross-influencing effects between advice items.

The original survey was advertised with a blog posting on the Google Online Security Blog [28]. Despite the support of the original authors, it was not possible to recruit developers the same way.

So instead we recruited experts through social media and mailing lists. We announced the survey link with a short advertising statement on Twitter[2] , asked selected professional contacts (e.g., the original authors) to repost or share the advertising; and also announced the study, together with a link to the tweet, on a hacking and security community mailing list. All in all, the tweet was retweeted 28 times and received 5,540 impressions, according to Twitter's analytics tool. In addition, the survey link was shared in the following reddit communities: r/Defcon, r/cybersecurity, r/netsecstudents, r/netsec, r/sysadmin, r/SampleSize, r/computerscience, r/information_Security, r/privacy.

## 4 Results

Of the 300 end user surveys that were completed, 12 participants got more than one of three quality assurance questions wrong and were, therefore, excluded from further analysis. This is the same procedure used in the original work. Our final sample thus consisted of 288 participants.

The collected demographic data is displayed in Table 1. The sample contained 48% female participants and was relatively young, with almost 80% of participants being younger than 45 years old. A little more than half have at least a bachelor's degree, and the majority, at 66%, reported an employment status of full-time employee. In comparison, the original study's sample had 40% female respondents, and 88% of the participants were younger than 45 years old. In the original study, 47% of the participants held a bachelor's degree or higher. In the original study, 47% of participants were from the US, data for EU-located participants was not given. In our sample, 70.4% of participants were from the EU and 26.8% were from the US.

The expert survey was conducted between June and November 2018. We recruited 75 expert participants online using our A/B testing design, 44 expert participants for survey form A (with compound questions), and 31 participants for survey form B (without compound questions). Participants were allowed one mistake regarding the three attention checks in the survey, as was done in the original study. We also excluded one participant who clearly gave nonsensical answers.

One prominent difference between our expert sample and the original expert sample is that our experts had less experience. The original study required experts to have at least five years of work or study experience in IT security or a related field. Only 59 participants fulfilled this requirement in our set, so we lowered this requirement to one year. We will discuss this in more detail in the limitations section.

[1] "For each of the following pieces of advice, please rate on a scale from 1 to 5 how good (in terms of both EFFECTIVE at keeping the user secure, as well as REALISTIC that the user can follow it) you think they are at protecting a non-tech-savvy user's security online."

[2] "Dear #security experts, I'm conducting a study about security advice targeted at non-technical users and need your help. Please participate in this 10-Minute survey: https://studyportal-bonn.de I appreciate RTs and (cross-platform) shares. Questions? DM or busse@cs.uni-bonn.de" (https://twitter.com/kb_usec/status/1047080662312898560)

| Item | NE | | E | |
|---|---|---|---|---|
| Female | 137 | 47.6% | 7 | 9.3% |
| Male | 150 | 52.1% | 59 | 78.7% |
| Transgender | 1 | 0.4% | 2 | 2.7% |
| No Answer | 0 | 0% | 7 | 9.3% |
| 18 - 24 | 25 | 8.7% | 3 | 4% |
| 25 - 34 | 130 | 45.1% | 30 | 40% |
| 35 - 44 | 72 | 25% | 26 | 34.7% |
| 45 - 54 | 39 | 13.5% | 9 | 12% |
| 55 - 64 | 16 | 5.6% | 2 | 2.7% |
| 65 or older | 6 | 2.1% | 0 | 0% |
| No answer | 0 | 0% | 5 | 6.7% |
| Professional Doctorate | 5 | 1.7% | 3 | 4% |
| Doctoral Degree | 3 | 1% | 6 | 8% |
| Master | 28 | 9.7% | 29 | 38.7% |
| Bachelor | 114 | 39.6% | 18 | 24% |
| Associates Degree | 38 | 13.2% | 3 | 4% |
| Some college, no degree | 45 | 15.6% | 4 | 5.3% |
| Technical/Trade School | 13 | 4.51% | 2 | 2.7% |
| Regular HS Diploma | 32 | 11.11% | 0 | 0% |
| GED or alternative | 5 | 1.74% | 0 | 0% |
| Some high school | 2 | 0.69% | 0 | 0% |
| Other | 0 | 0% | 4 | 5.3% |
| No answer | 3 | 1.04% | 6 | 8% |
| Employed full-time | 190 | 65.97% | | |
| Employed part-time | 26 | 28.26% | | |
| Self-employed | 36 | 12.50% | | |
| Homemaker | 16 | 5.56% | | |
| Retired | 6 | 2.08% | | |
| Student - Undergrad | 6 | 2.08% | | |
| Student - Doctoral | 2 | 0.69% | | |
| Looking for work | 9 | 3.13% | | |
| Other | 2 | 0.69% | | |
| Industry | | | 38 | 50.7% |
| University | | | 16 | 21.3% |
| Corporate research lab | | | 7 | 9.3% |
| Government | | | 1 | 1.3% |
| Self-employed | | | 2 | 2.7% |
| Other | | | 9 | 2.7% |
| No answer | | | 2 | 2.7% |
| 1-5 years of security exp. | | | 16 | 21.3% |
| 5-10 years of sec. exp. | | | 18 | 24.0% |
| 10-15 years of sec. exp. | | | 20 | 26.7% |
| 15+ years of sec. exp. | | | 21 | 28.0% |

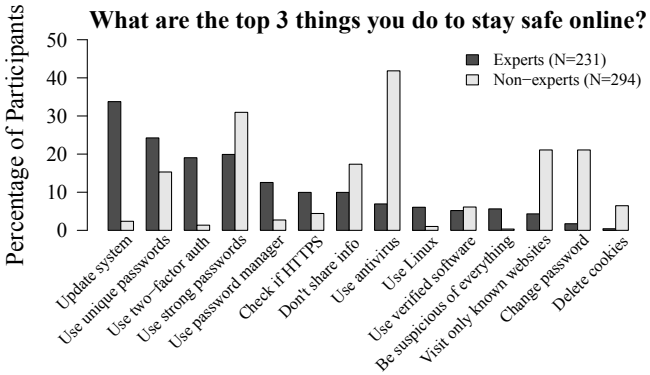
Table 1: Demographic information for expert (E, $n = 75$) and non-expert (NE, $n = 288$) survey participants.

The $p$ values we report refer to chi-squared tests or, where not enough data in all categories was available, Fisher's exact test. Dependent on the original authors' approach, we applied the Holm–Bonferroni correction in R for all the tests conducted. To further illustrate our results, we utilized participants' comments provided by the optional clarification questions and "other, please specify" options of the survey.

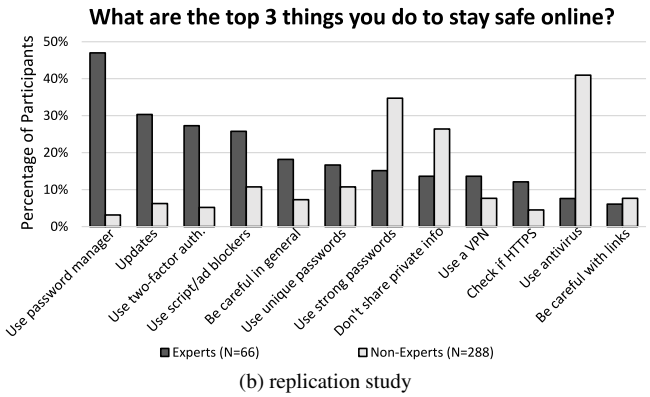
## 4.1 Differences between Experts and Non-Experts

For this section, we focus on experts and non-experts to follow the approach of the original work. Experts A and B were combined in behavior-related questions since these questions were identical, but split when advice rating was considered.

The first question asked about the top three things participants do to protect their security online. The comparison of the answers is displayed in Figure 1. In accordance with the original work, we only considered items mentioned by at least 5% of the participants in each group.



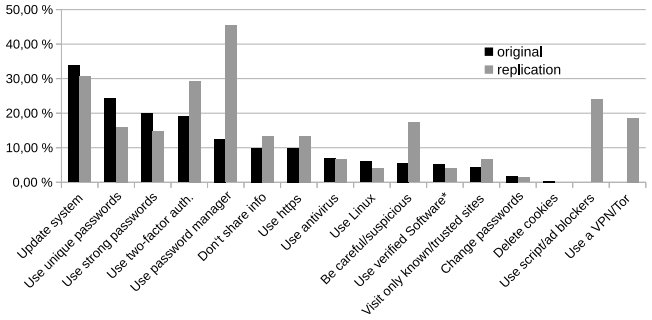

(a) original study



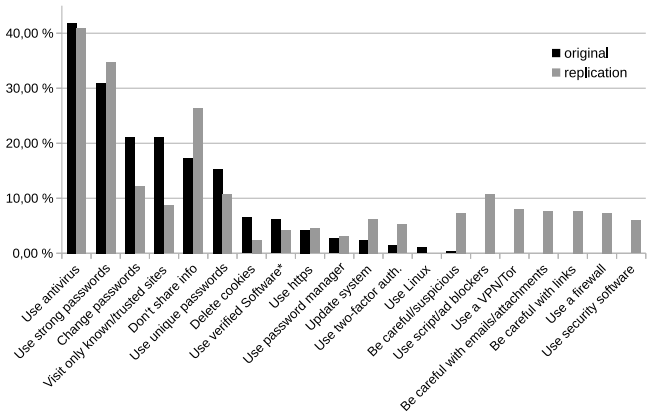

(b) replication study

Figure 1: Security measures mentioned by at least 5% of each group

While most experts rely on a password manager (45%) and updates (31%) as well as two-factor authentication (29%) to stay safe, non-experts count on the usage of antivirus software

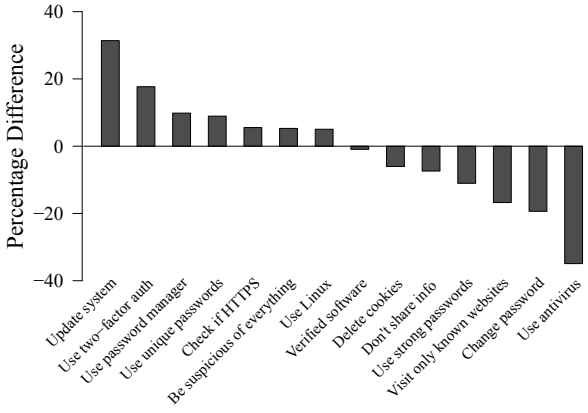



(a) Expert Comparison

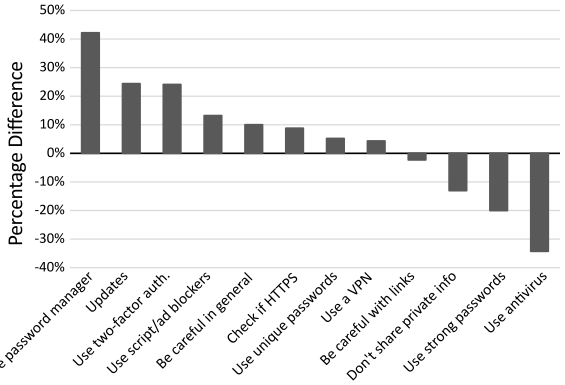



(b) Non-Expert Comparison

Figure 2: Answer comparison for the question "What are the top 3 things you do to stay safe online?" between the original study and our replication. Missing values for original data were mentioned by less than five percent of expert participants. (*) We aligned the original authors' code with our code "be careful with downloads".




(a) original study




(b) replication study

Figure 3: Percentage difference of security practices mentioned by experts and non-experts as answer to the "things-you-do" question. Security measures with a positive percentage difference were mentioned more by experts than non-experts; a negative percentage difference indicates topics mentioned more by non-experts.

(41%), strong passwords (35%), and not sharing personal information (26%).

In comparison with the original study, the most common security practice mentioned by experts has shifted. Instead of updating regularly, the use of a password manager was now the most frequently mentioned habit among our experts. The use of unique passwords, which was the original study's second most common practice, ranked sixth in our sample. Since the use of password managers usually includes the use of unique passwords, these two are linked. The adoption of two-factor authentication was unchanged, in position three.

Overall, there were four new practices frequently mentioned: using ad and/or script blockers, being careful in general as well as when following links, and using VPNs. In contrast, the once common practices of using Linux, using verified software, changing passwords regularly, and manually deleting cookies were not present in our sample. The replacement of "carefulness" with "practicing suspicion," however, might have been a product of different coding approaches.

The percentage differences between the groups of experts and non-experts are displayed in Figure 3. The practices mentioned least by non-experts relative to experts were: (1) use a password manager (42%), (2) keep your system up-to-date (24%), and (3) use two-factor authentication (24%). While the rankings of these three pieces of advice have shifted a bit (password managers climbed from difference position three to one), we still see the same overall trend as in 2014.

### 4.1.1 Software and OS Updates

As in the original study, we differentiated between operating system and application updates. In the question block about behavior with personal devices, we asked "How soon after you discover that a new version of your operating system (OS) software is available do you (or somebody else managing your computer) install it?" We saw that exactly half of all experts as well as non-experts reported installing their updates either *automatically* or *immediately* after they become available (cf.

| Reported Behavior | $\chi^2$ | $p$ |
|---|---|---|
| How soon do you install updates? | 7.95 | $< 0.001$ |
| Do you use antivirus software? | 77.43 | $< 0.001$ |
| Do you use two-factor authentication? | 23.41 | $< 0.001$ |
| Do you remember your passwords? | 35.43 | $< 0.001$ |
| Do you write down your passwords? | 20.03 | $< 0.001$ |
| Do you save your passwords in a file? | 1.79 | 0.651 |
| Do you use a password manager? | 55.59 | $< 0.001$ |
| Do you reuse passwords? | 21.43 | $< 0.001$ |
| Do you look at the URL bar? | 22.28 | 0.001 |
| Do you check if HTTPS? | 5.48 | $< 0.001$ |
| Do you visit websites you haven't heard of? | 48.16 | $< 0.001$ |
| Do you enter your PW on links in emails? | 63.95 | $< 0.001$ |
| Do you open emails from unknown? | 91.67 | $< 0.001$ |
| Do you click on links from unknown? | 16.52 | 0.013 |

Table 2: Comparing expert and non-expert reports on their security behavior. $N_e = 74, N_n = 282$ for the first two questions, otherwise $N_e = 75, N_n = 288$. Degrees of Freedom: 4 for the first, 1 for the second and third question, 3 otherwise. Fisher's Exact test instead of Pearson's Chi-Squared was used to calculate $p$ whenever not enough data was available in any category.

Figure 4). However, we can see that if compared to the findings of the original study, where 64% of experts and only 38% of non-experts installed their updates either automatically or immediately, fewer experts but more non-experts are reporting this behavior in our replication. While the numbers are closer together, the differences between the groups are still statistically significant ($\chi^2(4, N_e = 74, N_n = 282) = 7.95, p < 0.001$, cf. Table 2). This could be an artifact of widespread operating systems that employ automatic updates per default, as for example Windows 10 does.

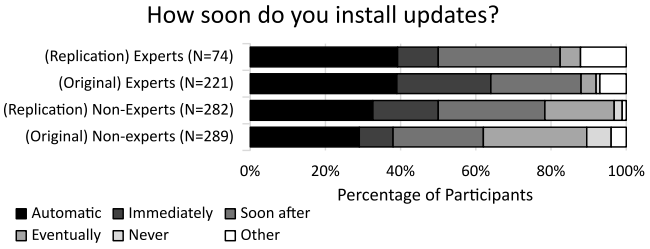



Figure 4: Answer distributions for the question "How soon after you discover that a new version of your operating system (OS) software is available do you (or somebody else managing your computer) install it? Examples of operating systems include Windows, Mac OS, and Linux.".

Among the pieces of advice, we had the statements "turn on automatic updates," "install OS updates," and "update applications." In all three cases of update-related advice, less than 50% of non-experts rated the advice very effective, yet around 60% said they were very likely to follow it. Especially for the advice regarding application updates, we found a strong discrepancy within our A/B testing setup. More about this is reported in Section 4.2

#### 4.1.2 Antivirus and Protection Software

Using antivirus software is still the security practice with the biggest difference in number of mentions between end users and experts (cf. Figure 3). As Figure 1 illustrates, 41% of non-experts and only 7% of experts stated that using antivirus software is one of the top three things they do to protect their security online. This coincides with the findings of the multiple-choice questions on security-related behavior in the second part of the survey, where twice as many non-experts as experts ($E = 82\%$ vs. $NE = 41\%$) reported using antivirus software on their personal computers. As shown in Table 2, this difference is statistically significant ($\chi^2(1, N_e = 74, N_n = 282) = 77.43, p < 0.001$).

Several experts stated that the perceived usefulness of antivirus software might be higher than the actual usefulness. One expert stated, "*I think antivirus software creates more problems than it solves (including the feeling of being safe).*" Some experts strongly suggested caution when dealing with antivirus software. One expert participant commented, "*Antivirus software often is snake-oil and detects only old viruses, but prevents users from these viruses. Also, they often implement suspicious features like breaking https without being clear to the end user about it.*"

Non-experts were asked to use a five-point Likert scale to rate how effective they see the security advice of using antivirus software: 63% rated it *very effective* and 19% rated it *effective*.

When asked how likely they would be to follow this advice if they heard that using antivirus software was effective, 73% of non-experts said they would be *very likely* to follow this advice, and 9% said they *likely* would. This strong acceptance of antivirus software is mirrored by the comments and feedback provided by non-experts.

A new type of security advice that emerged in the things-you-do question was the use of ad and/or script blockers. A proportion of 24% of experts and 11% of non-experts mentioned this security practice as one of their personal top three (cf. Figure 1).

#### 4.1.3 Password Management

In many cases, both experts and non-experts cited password-related practices as an answer to the question "What are the top three things you do to protect your security online?" Using strong and unique passwords were frequently mentioned strategies by both groups. Where experts spoke more of having unique passwords than non-experts ($E = 16\%$ vs. $NE = 11\%$), using strong passwords was reported twice as

often by non-experts than experts ($NE = 35\%$ vs. $E = 15\%$). While the practice of having unique passwords was mentioned less frequently than in the original data set (cf. Figure 2), having strong passwords was slightly less frequently mentioned by experts (then 20%, now 15%), but slightly more frequently mentioned by non-experts (then 31%, now 35%).

Similarly, experts named using a password manager substantially more often than non-experts ($E = 45\%$ vs. $NE = 3\%$), but almost did not mention changing passwords frequently (1% experts vs. 12% non-experts). Changing passwords is still not very prominent for experts (then 2%, now 1%), and has decreased in mentions by non-experts, as well (then 21%, now 15%; cf. Figure 2).

Likewise, experts mentioned the use of two-factor authentication more than five times as much as non-experts ($E = 29\%$ vs. $NE = 5\%$). This practice has gained in prominence for both experts (then 19%, now 29%) and non-experts (then 1%, now 5%). This could be partially attributed to the fact that more services now offer two-factor authentication than in 2014.

The most common answer of experts to the things-you-do question was "using a password manager" ($E = 45\%$), in contrast to a very small group of non-experts ($NE = 3\%$). In comparison with the original study, the mention of password managers by experts had more than tripled, from 13% to 45%. This difference is in line with the fact that twice as many experts as non-experts reported using a password manager for at least some of their accounts ($E = 83\%$ vs. $NE = 40\%$, $\chi^2(3, N_e = 75, N_n = 288) = 55.60$, $p < 0.001$). One expert commented, *"Using a proper password manager is the best solution. In the end, it is about using different passwords for different accounts."*

Writing down passwords was seen by some experts as a user-friendly compromise to a password manager. One expert said, *"[The advice to use] different passwords is effective, but can be difficult for users if they don't use a password manager. Writing passwords down isn't really bad, as long as the paper is kept secure. This is basically just an offline password manager."*

While the advice to "write down passwords on paper" and "save passwords in a file" were rated poorly by non-experts for both effectiveness and the likelihood that they would follow the advice if they heard it was secure, especially the practice of writing down passwords on paper, was rather common among our participants. As can be seen in Figure 5, 45% of non-experts reported writing down passwords for at least some of their accounts (vs. 33% of experts, $\chi^2(3, N_e = 75, N_n = 288) = 20.02$, $p < 0.001$). Almost all experts commented on the importance of storing the paper securely.

Also shown in Figure 5, six times more non-experts than experts remember all of their passwords (36% non-experts vs. 5% experts, $\chi^2(3, N_e = 75, N_n = 288) = 35.42$, $p < 0.001$). These numbers have decreased in comparison to the original study, where 17% of experts and 52% of non-experts cited being able to remember all of their passwords.

In addition, seven times more non-experts than experts stated that they reuse passwords for most or all of their accounts (23% of non-experts vs. 3% of experts, $\chi^2(3, N_e = 75, N_n = 288) = 21.43$, $p < 0.001$). While the proportion of end users who employ this practice rose slightly in comparison with the original study (19%), the rate among experts stayed about the same (3%).

#### 4.1.4 Mindfulness

Among the remaining pieces of advice, the ones about checking the URL bar when browsing and looking for HTTPS connections are most interesting in comparison to the original study, since there have been major changes in the SSL/TLS certificate ecosystem within the last few years.

The rise of Let's Encrypt and automated certificate issuance and renewal have greatly increased the level of TLS-encrypted web traffic [2]. In consequence, HTTPS has become more widespread, but the indication about whether a site should be trusted because it features HTTPS has been weakened, since even phishing websites often come with security certificates [33].

When asked about the advice to check if the website they're visiting uses HTTPS, 54% of non-experts rated it very effective, and 61% considered themselves very likely to follow that advice. In comparison, the original data featured a proportion of 60% of non-experts rating this advice as *very effective*, and 50% saying they would likely follow it.

To put this in context, we asked all participants whether they practice checking for HTTPS while surfing. The portion of experts who *often* do so decreased from 82% in the original study to 73% in our replication. The portion of non-experts increased from 36% in the original study to 47% in our replication.

Regarding the more general question about checking the URL bar when visiting a website, 76% of experts and 60% of non-experts said they often look at the URL bar (original study: 86% and 59%). Some experts emphasized that it is not only important to look at the URL bar, but also to be aware of the specific information it displays. For example, one expert said, *"Watch out for correct URLs, valid SSL certificates, and enabled encryption (HTTPS) if sensitive information is requested."*

The question whether a participant enters their passwords on websites after they click on a link in an email is the only behavior question for which the chi-squared test for expert and non-expert answers yielded a different result than in the original study. While Ion et al. found no significant difference between the groups, our samples showed a large effect size ($\chi^2(3, N_e = 75, N_n = 288) = 63.95$, $p < 0.001$). This results from a large proportion of expert users choosing the *Other* option to further explain their behavior in that case. While some experts stated in the comments that they generally do

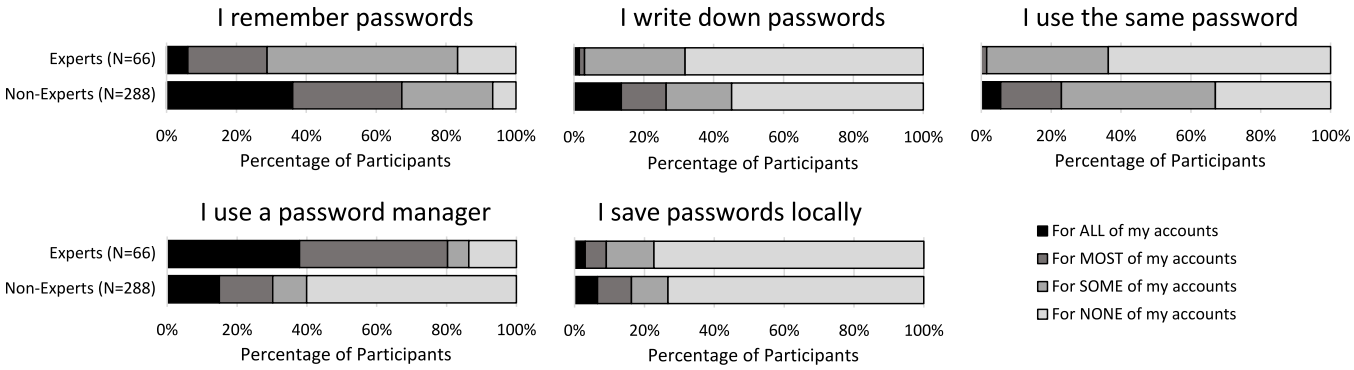



Figure 5: Non-expert habits regarding password management from our replication study.

not click on links in emails, another proportion of experts further differentiated, making comments such as, "*It depends. Am I expecting that email, is it from a reputable source, and does the URL match what I expect? Then yes; otherwise no.*" When excluding the *Other* option, the test results align again with the original study ($p = 0.63$ after correction).

## 4.2 Compound Question Results

As described in section 3.2, half the experts received the original survey with the compound questions (Group A) and for the other half we split up the goodness rating into effectiveness and realism (Group B). In the following, we compared the ratings of the split questions to those of the original compound questions.

In Figure 6, we look at the distribution of ratings given by experts A and B. Some pieces of advice, like installing OS updates, were rated very "effective" as well as very "realistic" by both expert groups. In the following, we will focus on the cases in which a piece of advice did not receive high scores in all cases, especially in terms of realism.

For example, not opening email attachments from unknown senders was rated positive in terms of goodness and effectiveness by experts A and experts B (64% *very good* and 16% *good* and 58% *very effective* and 35% *effective, respectively*). However, the *realistic* rating given by experts B peaks at a Likert score of 3, with 35%. Only 19% of experts B said this advice was very *realistic*, and 6% said it is *not realistic at all* (cf. Figure 7).

A piece of advice was classified as *good* and *effective* if a rating of 4 or better was present. We are most interested in those cases where this condition was met as well as having a realism rating of less than 4. As depicted in Table 3, this applies for eight pieces of advice.

We can group these pieces of advice in four categories.

Using unique and strong passwords as well as using a password manager all relate to *Password Security*. The advice to adopt *Two-Factor Authentication* stands on its own. Being suspicious of links, not entering passwords after having clicked on a link in an email, and not opening attachments can be grouped as *Links and Attachments*. The last piece of (controversial) advice, *Updating Applications* regularly, again stands on its own.

# 5 Discussion

In the following, we will discuss the popularity of selected findings and advice.

## 5.1 Advice Rating

While in the original study, the advice to regularly update showed the greatest difference between expert recommendations and non-expert usage, we found that using a password manager is now the piece of advice with the biggest gap between experts and end-users. Microsoft's shift toward mandatory automatic updates in Windows 10 might be the cause of this change. Because the operating system now takes care of keeping the system up to date, and thus secure, experts might not regard this advice to be as urgent as they did four years ago [22].

Password managers have the potential to solve the usability issue of passwords. Additionally, password managers might be a currently trending topic, which is reflected in the popularity of this practice as the single most frequently suggested piece of security behavior reported by experts (cf. Figure 1).

Installing and using antivirus software was the most frequently cited security measure by non-expert users in both the original study and our study. While antivirus software doesn't offer reliable protection against new and modified types of malware, the presence in advertising, as well as easy setup procedures, might have led to its unbroken popularity.

The advice to not share private information has become more important to both expert and non-expert users. However, one could argue that unconsciously shared information might, indeed, be more dangerous for users, whether it is conversation metadata [18, 31], tracking networks [1], or behavioral data like smartphone usage habits [19].

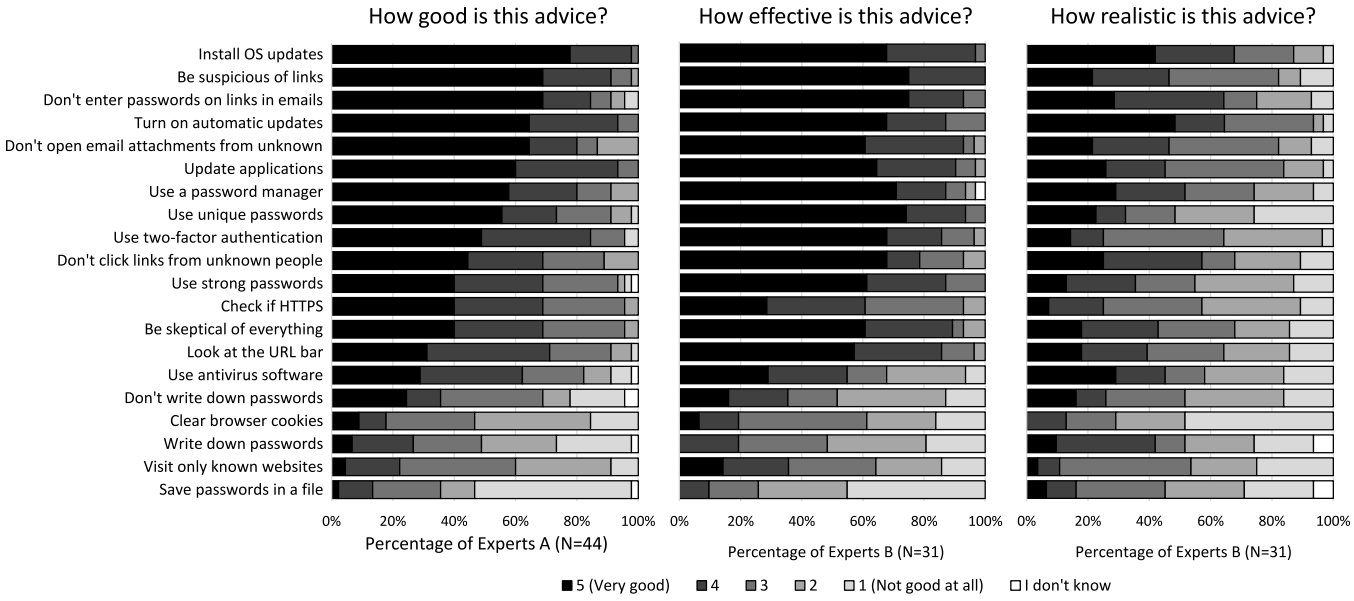



Figure 6: Side-by-side comparison of rating distributions in our replication study, showing from left to right: goodness ratings by experts A, efficiency ratings by experts B and realism ratings by experts B. The twenty pieces of advice are sorted by goodness ratings.

| **Advice** | $\delta_\mu$ | $\mu_e$ | $\mu_r$ | $\sigma_e$ | $\sigma_r$ | $\delta_m$ | $m_e$ | $m_r$ |
|---|---|---|---|---|---|---|---|---|
| Use unique passwords | 1.90 | 4.68 | 2.77 | 0.60 | 1.52 | 3 | 5 | 2 |
| Use strong passwords | 1.58 | 4.48 | 2.90 | 0.72 | 1.27 | 2 | 5 | 3 |
| Use two-factor authentication | 1.55 | 4.52 | 2.97 | 0.81 | 1.19 | 2 | 5 | 3 |
| Be suspicious of links | 1.35 | 4.71 | 3.35 | 0.46 | 1.28 | 2 | 5 | 3 |
| Use a password manager | 1.16 | 4.6 | 3.48 | 0.77 | 1.29 | 1 | 5 | 4 |
| Don't open email attachments | 1.16 | 4.48 | 3.32 | 0.72 | 1.17 | 2 | 5 | 3 |
| Don't enter PW on links in emails | 1.13 | 4.68 | 3.55 | 0.60 | 1.34 | 1 | 5 | 4 |
| Update applications | 1 | 4.51 | 3.52 | 0.77 | 1.12 | 2 | 5 | 3 |

Table 3: Pieces of advice that were received a mean effectiveness rating ($\mu_e$) of at least 4, and a mean realism rating ($\mu_r$) of less than 4, ordered by decreasing difference $\delta_\mu$. Also shown are standard deviations for effectiveness and realism ratings as well as medians and their difference.

## 5.2 New Advice

When looking at the free text answers for personal top three security practices, we found four new items within the top 18 most frequently mentioned statements: using script and/or ad blockers, being careful when online, using a VPN, and being careful when interacting with links (cf. Figure 1). In addition, five additional practices made it just beyond the 5% threshold: only visiting known or trusted websites, using incognito browsing, employing virtual machines, compartmentalizing systems for different tasks or levels of security, using a firewall, and employing security software in general. For the sake of brevity, we excluded these five practices from our further discussion.

While the more general advice of being careful might have arisen from different coding approaches between the original study and our replication, the other two pieces of advice suggest new developments.

Internet advertising has become more aggressive, invasive, and risky over the last few years [6], and blocking extensions are a powerful tool to combat this. In addition to the rise of this security practice, which 24% of the expert participants and 11% of the non-experts employ, the practice of manually deleting cookies was not included in the list anymore. This might be a replacement process, since many blocking tools also go after tracking cookies.

Using a VPN was a common response to the things-you-do question, but unfortunately, none of our participants elaborated on the meaning of this short statement. It is unclear exactly what kind of VPN participants were referring to. Just as Ferguson and Huston discovered two decades ago, "*VPN*

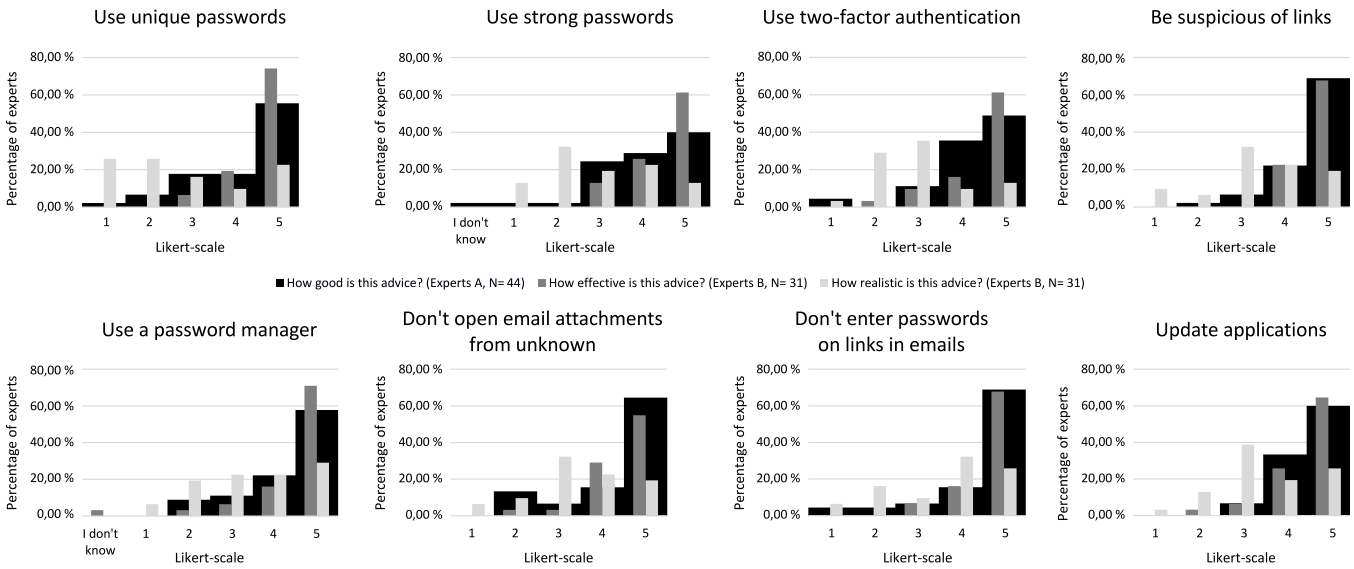



Figure 7: A/B comparison of advice rating from our replication study for pieces of advice identified as effective, but unrealistic. Descriptive statistics can be found in Table 3.

[has been and still is an almost] *recklessly used*" [12] collective term to describe various technologies and applications. VPNs and onion routing services such as Tor are effective tools for circumventing regional (e.g., governmental) censorship or content restrictions. However, using a VPN entails placing trust in its provider, which is a thing that users often overlook [14, 32].

## 5.3 Fields of Action

The pieces of advice that experts rated as very effective, but not very realistic for a user to follow, highlight areas where more research or better technical solutions are needed (cf. Table 3 and Figure 7). We identified four key fields of action; namely, password security, two-factor authentication, links and attachments, and application updates.

It is striking that these areas of advice are very similar to the advice not followed by users in the original study (cf. Section 2.1): recommending frequent system updates has been replaced with regular application updates, while using a password manager and enabling two-factor authentication have stayed the same.

### 5.3.1 Password Security

The advice ratings on unique and strong passwords indicate strongly that passwords are still an issue. The fact that the advice about adopting password managers also has a large delta between average effectiveness and realism ratings suggests that password managers are not yet fit for general adoption. Password managers should be approachable, easy to set up, and well-integrated into the operating system, without causing new security risks [8, 11].

However, even among experts, the use of password managers is not without drawbacks. One expert acknowledged a potentially "steep learning curve for non-tech-savvy users," while an end user stated that *"Storing passwords digitally and/or trusting a company to protect your data seems counterproductive."*

### 5.3.2 Two-Factor Authentication

Aside from the use of password managers, the adoption of two-factor authentication (2FA) is another relatively easy way to greatly increase account security. However, our expert group regarded this advice as not very realistic to be followed, while still acknowledging its effectiveness (cf. Table 3).

In general, more services need to support the setup of a second factor, since approximately 76% of websites do not offer users a full set of 2FA options [16]. Additionally, finding ways to increase user adoption of 2FA for accounts is a task for future research [3].

### 5.3.3 Links and Attachments

Three statements in our list of controversially rated advice related to links and attachments, specifically, being suspicious of links, not entering passwords on links received in emails, and not opening email attachments.

While the experts might have rated it as not very realistic, since opening attachments and following links is part of daily internet life, the risks arising from well-crafted phishing or malware emails should not to be neglected. A prominent example from recent years is the rise of ransomware, like wannacry [29].

Protecting against these types of threats purely from the technical side is rather difficult since they usually come with a measure of social engineering. Phishing URLs increasingly make use of invisible Unicode characters or identical-looking symbols from non-Latin alphabets [35].

One possible solution for preventing malware infection after opening an email or its attachments could be sandboxing technology. All attachments and links would be opened in an isolated, secure environment that doesn't harm the actual system.

#### 5.3.4 Application Updates

Last but not least, our results suggest further research in the direction of update managers that not only reliably perform their task of keeping the system and its applications up to date, but also communicate clearly what updates include which features and fixes and that schedule their work intelligently without interrupting or hindering the user.

The need for a centralized, system-level update tool that takes care of application updates was already expressed by Ion, Reeder, and Consolvo in 2015 and recently confirmed by Mathur et al. [21]. Since then, some applications have started to implement their own more or less automatic update tools, while a centralized tool is not on the horizon. Microsoft tried establishing their own Windows Store as an app store-like entity with an integrated application updater, but adoption rates are still low.

## 6 Limitations

In the following, we will outline the limitations of our study to facilitate putting this work into context.

Because we could not recruit via the same channel as the original authors, our expert sample is drawn from a different population. Thus, there are two variables that are different, time and population from which our experts were recruited. For that reason, our results can be seen as extending the original results, but cannot be used to state that the effects are attributable to the intervening time or due to different populations.

In particular, we decided to include security experts with 1-5 years of experience in security or a related field, while the original study only considered participants with at least five years of experience as experts. Table 1 shows that participant distribution is almost equal between all age brackets. Since we saw no difference between experts with 1-5 years of experience and those with 5+ years of experience, we decided to include them to increase our overall sample size.

As for recruiting non-experts, we had to follow the same channel as the original work and thus suffer from the same limitations. While Amazon MTurk is heavily used for usable security and human–computer interaction studies, the population there tends to be younger, more female, and more tech-savvy than the general US population [7, 25, 30].

All data we collected were self-reported. It is known that people tend to put themselves in a better light in such situations; therefore, the adoption rates or likeliness of following a certain piece of advice are possibly skewed [27].

## 7 Conclusions

In this paper, we replicated a 2015 study by Ion, Reeder, and Consolvo examining expert and non-expert security habits and corresponding advice. While our general findings relate with the original work, we could identify some new trends, like the use of script and ad blocking software.

In addition, we identified an issue in the original study design and improved upon it. Our results identify critical areas of effective but unrealistic practices that could be improved upon by the research and practitioner communities. Most of these practices (password security, 2FA, securely handling links and attachments from emails, and centralizing application updates) were already present as emerging topics in the 2015 study. This shows that the usable security community has not succeeded in solving these grave issues and clearly outlines the need for future action in researching and developing new or better security tools that non-experts can adopt and use.

## 8 Acknowledgments

We thank Iulia Ion, Rob Reeder, and Sunny Consolvo for their cooperation and support of our replication efforts. Furthermore, we thank all the people who helped in distributing and sharing the expert survey. Many thanks go to Maximilian Häring, Emanuel von Zezschwitz, and Christian Tiefenau for feedback on the paper. We also thank all our study participants as well as the anonymous reviewers.

This work was partially funded by the ERC Grant 678341: Frontiers of Usable Security.

## References

[1] G. Acar, C. Eubank, S. Englehardt, M. Juarez, A. Narayanan, and C. Diaz. The web never forgets: Persistent tracking mechanisms in the wild. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, pages 674–689. ACM, 2014.

[2] M. Aertsen, M. Korczyński, G. Moura, S. Tajalizadehkhoob, and J. van den Berg. No domain left behind: is let's encrypt democratizing encryption? In *Proceedings of the Applied Networking Research Workshop*, pages 48–54. ACM, 2017.

[3] Y. Albayram, M. M. H. Khan, and M. Fagan. A study on designing video tutorials for promoting security features: A case study in the context of two-factor authentication (2fa). *International Journal of Human–Computer Interaction*, 33(11):927–942, 2017.

[4] E. R. Babbie and L. Benaquisto. *Fundamentals of social research*. Cengage Learning, 2009.

[5] Z. Benenson, G. Lenzini, D. Oliveira, S. Parkin, and S. Uebelacker. Maybe poor johnny really cannot encrypt: The case for a complexity theory for usable security. In *Proceedings of the 2015 New Security Paradigms Workshop*, pages 85–99. ACM, 2015.

[6] R. Benes. Five Charts: Why Users Are Fed Up with Digital Ads. https://www.emarketer.com/content/five-charts-users-are-fed-up-with-digital-ads, 2018. last accessed on 2019-02-20.

[7] M. Buhrmester, T. Kwang, and S. D. Gosling. Amazon's mechanical turk. *Perspectives on Psychological Science*, 6(1):3–5, 2011. PMID: 26162106.

[8] S. Chiasson, P. C. van Oorschot, and R. Biddle. A usability study and critique of two password managers. In *USENIX Security Symposium*, pages 1–16, 2006.

[9] S. H. Drew. Drew: Tips on creating passwords to protect your privacy. https://www.birminghamtimes.com/2018/11/drew-tips-on-creating-passwords-to-protect-your-privacy/, November 2018. last accessed on 2018-12-09.

[10] M. Fagan and M. M. H. Khan. Why do they do what they do?: A study of what motivates users to (not) follow computer security advice. In *Twelfth symposium on usable privacy and security (SOUPS 2016)*, pages 59–75, 2016.

[11] S. Fahl, M. Harbach, M. Oltrogge, T. Muders, and M. Smith. Hey, you, get off of my clipboard. In *International Conference on Financial Cryptography and Data Security*, pages 144–161. Springer, 2013.

[12] P. Ferguson and G. Huston. What is a vpn? - part i. *The Internet Protocol Journal*, 1(1), 1998.

[13] M. Harbach, S. Fahl, and M. Smith. Who's afraid of which bad wolf? a survey of it security risk awareness. In *Computer Security Foundations Symposium (CSF), 2014 IEEE 27th*, pages 97–110. IEEE, 2014.

[14] M. Ikram, N. Vallina-Rodriguez, S. Seneviratne, M. A. Kaafar, and V. Paxson. An analysis of the privacy and security risks of android vpn permission-enabled apps. In *Proceedings of the 2016 Internet Measurement Conference*, IMC '16, pages 349–364, New York, NY, USA, 2016. ACM.

[15] I. Ion, R. Reeder, and S. Consolvo. "...No one Can Hack My Mind": Comparing Expert and Non-Expert Security Practices. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*, pages 327–346, Ottawa, 2015. USENIX Association.

[16] E. Katz. Dashlane research finds majority of two-factor authentication offerings fall short. https://blog.dashlane.com/2fa-rankings/, 2018. last accessed on 2018-12-19.

[17] A. Lenhart, M. Madden, S. Cortesi, U. Gasser, and A. Smith. Where teens seek online privacy advice. *Pew Research Center, Internet & Technology*, 2013.

[18] P. Leonard. Mandatory Internet Data Retention in Australia – Looking the horse in the mouth after it has bolted. Technical report, Gilbert & Tobin, 2015.

[19] Y. Li, W. Dai, Z. Ming, and M. Qiu. Privacy protection for preventing data over-collection in smart city. *IEEE Transactions on Computers*, 65(5):1339–1350, 2016.

[20] E. L. MacGeorge, B. Feng, and E. R. Thompson. "good" and "bad" advice. *Studies in applied interpersonal communication*, 145, 2008.

[21] A. Mathur, N. Malkin, M. Harbach, E. Peer, and S. Egelman. Quantifying users' beliefs about software updates. *Proceedings 2018 Workshop on Usable Security*, 2018.

[22] J. Morris, I. Becker, and S. Parkin. In Control with no Control: Perceptions and Reality of Windows 10 Home Edition Update Features. In *Workshop on Usable Security and Privacy (USEC)*, 2019.

[23] NCSA. The stay safe online blog. https://staysafeonline.org/blog_category/privacy/, July 2018. last accessed on 2018-12-09.

[24] E. M. Redmiles, S. Kross, and M. L. Mazurek. How i learned to be secure: a census-representative survey of security advice sources and behavior. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 666–677. ACM, 2016.

[25] E. M. Redmiles, S. Kross, and M. L. Mazurek. How well do my results generalize? comparing security and privacy survey results from mturk, web, and telephone samples. In *How Well Do My Results Generalize? Comparing Security and Privacy Survey Results from MTurk, Web, and Telephone Samples*, page 0. IEEE, 2019.

[26] E. M. Redmiles, A. R. Malone, and M. L. Mazurek. I think they're trying to tell me something: Advice sources and selection for digital security. In *Security and Privacy (SP), 2016 IEEE Symposium on*, pages 272–288. IEEE, 2016.

[27] E. M. Redmiles, Z. Zhu, S. Kross, D. Kuchhal, T. Dumitras, and M. L. Mazurek. Asking for a friend: Evaluating response biases in security user studies. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, pages 1238–1255. ACM, 2018.

[28] R. W. Reeder. If you cold tell a user three things to do to stay safe online, what would they be? https://security.googleblog.com/2014/03/if-you-could-tell-user-three-things-to.html, March 2014. last accessed on 2019-02-20.

[29] J.-L. Richet. Extortion on the internet: the rise of crypto-ransomware. *Harvard*, 2016.

[30] D. J. Simons and C. F. Chabris. Common (mis)beliefs about memory: A replication and comparison of telephone and Mechanical Turk survey methods". *PLOS ONE*, 7(12), 2012.

[31] C. Simpson. Data Mining of Telecom Metadata is "More Dangerous than Intercepting Conversations". https://newsmonitors.blog/2018/04/19/data-mining-of-telecom-metadata-is-more-dangerous-than-intercepting-conversations/, 2018. last accessed on 2019-02-14.

[32] W. Strayer. Privacy issues in virtual private networks. *Computer Communications*, 27(6):517 – 521, 2004. Internet Performance and Control of Network Systems.

[33] E. Volkman. 49 percent of phishing sites now use https. https://info.phishlabs.com/blog/49-percent-of-phishing-sites-now-use-https, 2018. last accessed on 2019-02-13.

[34] R. Wash. Folk models of home computer security. In *Proceedings of the Sixth Symposium on Usable Privacy and Security*, page 11. ACM, 2010.

[35] X. Zheng. Phishing with Unicode Domains. https://www.xudongz.com/blog/2017/idn-phishing/, 2017. last accessed on 2019-02-14.

## A Surveys

All multiple-choice questions were single answer only. The questions were identical for the Expert A, Expert B, and Non-expert survey, unless otherwise stated. The questions marked "(Experts A only)", "(Experts B only)" or "(Non-experts only)" were asked in only one of the surveys.

- *(Experts A&B only)* What are the top 3 pieces of advice you would give to a non-tech-savvy user to protect their security online? *(open-ended)*
- What are the 3 most important things you do to protect your security online? *(open-ended)*
- How did you learn about the things you listed above? *(open-ended)*
- Do you use a laptop or desktop computer that you or your family owns (i.e., not provided by school or work)? *(multiple-choice)*
  - Yes
  - No
- When did you get that computer? *(multiple-choice)*
  - Less than 1 year ago
  - At least 1 but less than 2 years ago
  - At least 2 but less than 3 years ago
  - At least 3 but less than 5 years ago
  - 5 or more years ago
  - I don't know
- How soon after you discover that a new version of your operating system (OS) software is available do you (or somebody else managing your computer) install it? *(multiple-choice)*
  - OS updates are installed automatically
  - Immediately
  - Soon after
  - Eventually
  - OS updates are never installed
  - Other *(open-ended)*
- Do you use anti-virus software on that computer? *(multiple-choice)*
  - Yes
  - No
  - I don't know
  - Other *(open-ended)*
- Which anti-virus software do you use? *(open-ended)*
- How do you keep track of your passwords for your online accounts? *(grid question)*
  Answer options: For ALL of my accounts, For MOST of my accounts, For SOME of my accounts, For NONE of my accounts
  - Remember them
  - Write them down on paper
  - Save them in a local file on my computer

  - Have my password manager (e.g., 1Password, LastPass) remember them
  - Use the same password on multiple accounts
- If you use a password manager, which one do you use? *(open-ended)*
- (optional) What other things, if any, do you do to keep track of your passwords? *(open-ended)*
- Do you use two-factor authentication (e.g., 2-Step Verification) for at least one of your online accounts? *(multiple-choice)*
  - Yes
  - No
  - I don't know
  - Other *(open-ended)*
- Do you look at the URL bar to verify that you are visiting the website you intended to? *(multiple-choice)*
  - Yes, often
  - Yes, sometimes
  - Yes, rarely
  - No
  - I don't know
  - Other *(open-ended)*
- Google began in January 1996 as a research project. Its initial public offering took place on August 19, 2004. Did the initial public offering of Google take place in 1996? *(multiple-choice)*
  - Yes
  - No
  - Other *(open-ended)*
- Do you check if the website you're visiting uses HTTPS? *(multiple-choice)*
  - Yes, often
  - Yes, sometimes
  - Yes, rarely
  - No
  - I don't know
  - Other *(open-ended)*
- Do you visit websites you have not heard of before? *(multiple-choice)*
  - Yes, often
  - Yes, sometimes
  - Yes, rarely
  - No
  - I don't know
  - Other *(open-ended)*
- When you click on a link in an email and that link takes you to a website that asks for your password, do you enter it? *(multiple-choice)*
  - Yes, often
  - Yes, sometimes
  - Yes, rarely
  - No
  - I don't know
  - Other *(open-ended)*

  Do you open emails you receive from people or companies you don't know? *(multiple-choice)*
  - Yes, often
  - Yes, sometimes
  - Yes, rarely
  - No
  - I don't know
  - Other *(open-ended)*
- Do you click on links that people or companies you don't know send you? *(multiple-choice)*
  - Yes, often
  - Yes, sometimes
  - Yes, rarely
  - No
  - I don't know
  - Other *(open-ended)*
- *(Experts A only)* For each of the following pieces of advice, please rate on a scale from 1 to 5 how good (in terms of both EFFECTIVE at keeping the user secure, as well as REALISTIC that the user can follow it) you think they are at protecting a non-tech-savvy user's security online. *(grid question)*
  Scale: 5 (Very good), 4, 3, 2, 1 (Not at all), I don't know
  - Use anti-virus software
  - Install the latest operating system updates
  - Turn on automatic software updates
  - Update applications to the latest version
  - Clear your Web browser cookies

- *(Experts B only)* For each of the following pieces of advice, please rate on a scale from 1 to 5 how EFFECTIVE (at keeping the user secure) you think they are at protecting a non-tech-savvy user's security online. *(grid question)*
  Scale: 5 (Very good), 4, 3, 2, 1 (Not at all), I don't know
  - Use anti-virus software
  - Install the latest operating system updates
  - Turn on automatic software updates
  - Update applications to the latest version
  - Clear your Web browser cookies
- *(Non-experts only)* For each of the following pieces of advice, please rate on a scale from 1 to 5 how EFFECTIVE you think the advice would be at protecting your security online, IF YOU FOLLOWED IT. *(grid question)*
  Scale: 5 (Very effective), 4, 3, 2, 1 (Not at all), I don't know
  - Use anti-virus software
  - Install the latest operating system updates
  - Turn on automatic software updates
  - Update applications to the latest version
  - Clear your Web browser cookies
- *(Non-experts & Experts A only)*(optional) Please use this space to clarify any of the above. *(open-ended)*
- *(Experts B only)* For each of the following pieces of advice, please rate on a scale from 1 to 5 how REALISTIC (that the user can follow it) you think they are at protecting a non-tech-savvy user's security online. *(grid question)*
  Scale: 5 (Very good), 4, 3, 2, 1 (Not at all), I don't know
  - Use anti-virus software
  - Install the latest operating system updates
  - Turn on automatic software updates
  - Update applications to the latest version
  - Clear your Web browser cookies
- *(Non-experts only)* For each of the following pieces of advice, please rate on a scale from 1 to 5 how LIKELY YOU WOULD BE TO FOLLOW the advice, if you heard it would help protect your security online. *(grid question)*
  Scale: 5 (Very likely), 4, 3, 2, 1 (Not at all), I don't know
  - Use anti-virus software
  - Install the latest operating system updates
  - Turn on automatic software updates
  - Update applications to the latest version
  - Clear your Web browser cookies
- *(Non-experts & Experts B only)* (optional) Please use this space to clarify any of the above. *(open-ended)*
- *(Experts A only)* For each of the following pieces of advice, please rate on a scale from 1 to 5 how good (in terms of both EFFECTIVE at keeping the user secure, as well as REALISTIC that the user can follow it) you think they are at protecting a non-tech-savvy user's security online. *(grid question)*
  Scale: 5 (Very good), 4, 3, 2, 1 (Not at all), I don't know
  - Use different passwords for each account
  - Use passwords that are not easy to guess
  - Don't write down passwords on paper
  - Save your passwords in a local file on their computer
  - Use a password manager (e.g., 1Password, LastPass)
  - Write down passwords on paper
- *(Experts B only)* For each of the following pieces of advice, please rate on a scale from 1 to 5 how EFFECTIVE (at keeping the user secure) you think they are at protecting a non-tech-savvy user's security online. *(grid question)*
  Scale: 5 (Very good), 4, 3, 2, 1 (Not at all), I don't know
  - Use different passwords for each account
  - Use passwords that are not easy to guess
  - Don't write down passwords on paper
  - Save your passwords in a local file on their computer
  - Use a password manager (e.g., 1Password, LastPass)
  - Write down passwords on paper
- *(Non-experts only)* For each of the following pieces of advice, please rate on a scale from 1 to 5 how EFFECTIVE you think the advice would be at protecting your security online, IF YOU FOLLOWED IT. *(grid question)*
  Scale: 5 (Very effective), 4, 3, 2, 1 (Not at all), I don't know
  - Use different passwords for each account
  - Use passwords that are not easy to guess
  - Don't write down passwords on paper

  - Save your passwords in a local file on their computer
  - Use a password manager (e.g., 1Password, LastPass)
  - Write down passwords on paper

- *(Non-experts & Experts A only)* (optional) Please use this space to clarify any of the above. *(open-ended)*

- *(Experts B only)* For each of the following pieces of advice, please rate on a scale from 1 to 5 how REALISTIC (that the user can follow it) you think they are at protecting a non-tech-savvy user's security online. *(grid question)*
  Scale: 5 (Very good), 4, 3, 2, 1 (Not at all), I don't know
  - Use different passwords for each account
  - Use passwords that are not easy to guess
  - Don't write down passwords on paper
  - Save your passwords in a local file on their computer
  - Use a password manager (e.g., 1Password, LastPass)
  - Write down passwords on paper

- *(Non-experts only)* For each of the following pieces of advice, please rate on a scale from 1 to 5 how LIKELY YOU WOULD BE TO FOLLOW the advice, if you heard it would help protect your security online. *(grid question)*
  Scale: 5 (Very likely), 4, 3, 2, 1 (Not at all), I don't know
  - Use different passwords for each account
  - Use passwords that are not easy to guess
  - Don't write down passwords on paper
  - Save your passwords in a local file on their computer
  - Use a password manager (e.g., 1Password, LastPass)
  - Write down passwords on paper

- *(Non-experts & Experts B only)* (optional) Please use this space to clarify any of the above. *(open-ended)*

- *(Experts A only)* For each of the following pieces of advice, please rate on a scale from 1 to 5 how good (in terms of both EFFECTIVE at keeping the user secure, as well as REALISTIC that the user can follow it) you think they are at protecting a non-tech-savvy user's security online. *(grid question)*
  Scale: 5 (Very good), 4, 3, 2, 1 (Not at all), I don't know
  - Check if the website you're visiting uses HTTPS
  - Be skeptical of everything when online
  - Be suspicious of links received in emails or messages
  - Visit only websites you've heard of
  - Use two-factor authentication for your online accounts

- *(Experts B only)* For each of the following pieces of advice, please rate on a scale from 1 to 5 how EFFECTIVE (at keeping the user secure) you think they are at protecting a non-tech-savvy user's security online. *(grid question)*
  Scale: 5 (Very good), 4, 3, 2, 1 (Not at all), I don't know
  - Check if the website you're visiting uses HTTPS
  - Be skeptical of everything when online
  - Be suspicious of links received in emails or messages
  - Visit only websites you've heard of
  - Use two-factor authentication for your online accounts

- *(Non-experts only)* For each of the following pieces of advice, please rate on a scale from 1 to 5 how EFFECTIVE you think the advice would be at protecting your security online, IF YOU FOLLOWED IT. *(grid question)*
  Scale: 5 (Very effective), 4, 3, 2, 1 (Not at all), I don't know
  - Check if the website you're visiting uses HTTPS
  - Be skeptical of everything when online
  - Be suspicious of links received in emails or messages
  - Visit only websites you've heard of
  - Use two-factor authentication for your online accounts

- *(Non-experts & Experts A only)* (optional) Please use this space to clarify any of the above. *(open-ended)*

- *(Experts B only)* For each of the following pieces of advice, please rate on a scale from 1 to 5 how REALISTIC (that the user can follow it) you think they are at protecting a non-tech-savvy user's security online. *(grid question)*
  Scale: 5 (Very good), 4, 3, 2, 1 (Not at all), I don't know
  - Check if the website you're visiting uses HTTPS
  - Be skeptical of everything when online
  - Be suspicious of links received in emails or messages

  - Visit only websites you've heard of
  - Use two-factor authentication for your online accounts

- *(Non-experts only)* For each of the following pieces of advice, please rate on a scale from 1 to 5 how LIKELY YOU WOULD BE TO FOLLOW the advice, if you heard it would help protect your security online. *(grid question)*
  Scale: 5 (Very likely), 4, 3, 2, 1 (Not at all), I don't know
  - Check if the website you're visiting uses HTTPS
  - Be skeptical of everything when online
  - Be suspicious of links received in emails or messages
  - Visit only websites you've heard of
  - Use two-factor authentication for your online accounts

- *(Non-experts & Experts B only)* (optional) Please use this space to clarify any of the above. *(open-ended)*

- *(Experts only)* For each of the following pieces of advice, please rate on a scale from 1 to 5 how good (in terms of both EFFECTIVE at keeping the user secure, as well as REALISTIC that the user can follow it) you think they are at protecting a non-tech-savvy user's security online. *(grid question)*
  Scale: 5 (Very good), 4, 3, 2, 1 (Not at all), I don't know
  - Don't click on links that people or companies you don't know send you
  - Don't enter your password when you click on a link in an email and that link takes you to a website that asks for your password
  - Pay attention when taking online surveys. We appreciate your input. To let us know you're paying attention, select four for this response
  - Look at the URL bar to verify that you are visiting the website you intended to
  - Don't open email attachments from people or companies you don't know

- *(Experts B only)* For each of the following pieces of advice, please rate on a scale from 1 to 5 how EFFECTIVE (at keeping the user secure) you think they are at protecting a non-tech-savvy user's security online. *(grid question)*
  Scale: 5 (Very good), 4, 3, 2, 1 (Not at all), I don't know
  - Don't click on links that people or companies you don't know send you
  - Don't enter your password when you click on a link in an email and that link takes you to a website that asks for your password
  - Pay attention when taking online surveys. We appreciate your input. To let us know you're paying attention, select four for this response
  - Look at the URL bar to verify that you are visiting the website you intended to
  - Don't open email attachments from people or companies you don't know

- *(Non-experts only)* For each of the following pieces of advice, please rate on a scale from 1 to 5 how EFFECTIVE you think the advice would be at protecting your security online, IF YOU FOLLOWED IT. *(grid question)*
  Scale: 5 (Very effective), 4, 3, 2, 1 (Not at all), I don't know
  - Don't click on links that people or companies you don't know send you
  - Don't enter your password when you click on a link in an email and that link takes you to a website that asks for your password
  - Pay attention when taking online surveys. We appreciate your input. To let us know you're paying attention, select four for this response
  - Look at the URL bar to verify that you are visiting the website you intended to
  - Don't open email attachments from people or companies you don't know

- *(Non-experts & Experts A only)* (optional) Please use this space to clarify any of the above. *(open-ended)*

- *(Experts B only)* For each of the following pieces of advice, please rate on a scale from 1 to 5 how REALISTIC (that the user can follow it) you think they are at protecting a non-tech-savvy user's security online. *(grid question)*
  Scale: 5 (Very good), 4, 3, 2, 1 (Not at all), I don't know
  - Don't click on links that people or companies you don't know send you
  - Don't enter your password when you click on a link in an email and that link takes you to a website that asks for your password
  - Pay attention when taking online surveys. We appreciate your input. To let us know you're paying attention, select four for this response
  - Look at the URL bar to verify that you are visiting the website you intended to

  - Don't open email attachments from people or companies you don't know

- *(Non-experts only)* For each of the following pieces of advice, please rate on a scale from 1 to 5 how LIKELY YOU WOULD BE TO FOLLOW the advice, if you heard it would help protect your security online. *(grid question)*
  Scale: 5 (Very likely), 4, 3, 2, 1 (Not at all), I don't know
  - Don't click on links that people or companies you don't know send you
  - Don't enter your password when you click on a link in an email and that link takes you to a website that asks for your password
  - Pay attention when taking online surveys. We appreciate your input. To let us know you're paying attention, select four for this response
  - Look at the URL bar to verify that you are visiting the website you intended to
  - Don't open email attachments from people or companies you don't know
- *(Non-experts & Experts B only)* (optional) Please use this space to clarify any of the above. *(open-ended)*
- What is your gender? *(multiple-choice)*
  - Female
  - Male
  - Transgender
  - I prefer not to answer
  - Other *(open-ended)*
- What is your age? *(multiple-choice)*
  - 18-24 years old
  - 25-34
  - 35-44
  - 45-54
  - 55-64
  - 65 or older
  - I prefer not to answer
- What is the highest degree or level of school that you have completed? *(multiple-choice)*
  - Professional doctorate (for example, MD, JD, DDS, DVM, LLB)
  - Doctoral degree (for example, PhD, EdD)
  - Masters degree (for example, MS, MBA, MEng, MA, MEd, MSW)
  - Bachelor (for example, BS, BA; also German Berufsausbildung)
  - Associates Degree (or German Abitur)
  - Some college, no degree
  - Technical/Trade school
  - Regular High School Diploma (or German Realschulabschluss)
  - GED or alternative credential
  - Some High School (or German Hauptschulabschluss)
  - I prefer not to answer
  - Other *(open-ended)*
- *(Experts A&B only)* How many total years of experience do you have in computer security?
  'Experience' includes years at work or studying in a security-related field. *(multiple-choice)*
  - At least 1 but less than 5 years
  - At least 5 but less than 10 years
  - At least 10 but less than 15 years
  - 15 years or more
  - None
- *(Experts A&B only)* What is your current job role?
  For example, Network Security Engineer, Penetration Tester *(open-ended)*
  - Researcher
  - Principal Architect
  - IT Strategist
  - CEO
  - Manager
  - Security Engineer
  - Engineer
  - Other *(open-ended)*
- *(Experts A&B only)* Which of the following best characterizes your workplace? *(multiple-choice)*
  - University
  - Corporate research lab
  - Industry
  - Government
  - Self-employed
  - Other *(open-ended)*
- *(Experts A&B only)* In what country do you work? *(multiple-choice)*

  - Australia
  - Canada
  - Germany
  - India
  - United Kingdom
  - United States
  - Other *(open-ended)*
- *(Experts A&B only)* In what state do you work? (open-choice)
- *(Non-experts only)* Which describes your current employment status? *(multiple-choice)*
  - Employed full-time
  - Employed part-time
  - Self-employed
  - Care-provider
  - Homemaker
  - Retired
  - Student - Undergraduate
  - Student - Masters
  - Student - Doctoral
  - Looking for work / Unemployed
  - Other *(open-ended)*
- *(Non-experts only)* What is your occupation? *(open-ended)*
- *(Non-experts only)* What is your Mechanical Turk Worker ID? *(open-ended)*
- *(Experts A&B only)* Do you remember taking a survey with similar questions in the past (ca. 2014)?
  - Yes
  - No
- (Optional) Is there anything else you'd like to add or clarify? *(open-ended)*

# "*Something isn't secure, but I'm not sure how that translates into a problem*": Promoting autonomy by designing for understanding in Signal

Justin Wu
*Brigham Young University*

Cyrus Gatrell
*Brigham Young University*

Devon Howard
*Brigham Young University*

Jake Tyler
*Brigham Young University*

Elham Vaziripour
*Utah Valley University*

Kent Seamons
*Brigham Young University*

Daniel Zappala
*Brigham Young University*

## Abstract

Security designs that presume enacting secure behaviors to be beneficial in all circumstances discount the impact of response cost on users' lives and assume that all data is equally worth protecting. However, this has the effect of reducing user autonomy by diminishing the role personal values and priorities play in the decision-making process. In this study, we demonstrate an alternative approach that emphasizes users' comprehension over compliance, with the goal of helping users to make more informed decisions regarding their own security. To this end, we conducted a three-phase redesign of the warning notifications surrounding the authentication ceremony in Signal. Our results show how improved comprehension can be achieved while still promoting favorable privacy outcomes among users. Our experience reaffirms existing arguments that users should be empowered to make personal trade-offs between perceived risk and response cost. We also find that system trust is a major factor in users' interpretation of system determinations of risk, and that properly communicating risk requires an understanding of user perceptions of the larger security ecosystem in whole.

## 1 Introduction

The primary goal of usable security and privacy is to empower users to keep themselves safe from threats to their security or privacy. Their ability to do so is reliant on an accurate assessment of the existence and severity of a given risk, the set of available responses, and the cost of enacting those responses. Ideally, users would like to take action only when a threat has been realized *and* the negative consequences of that threat are severe enough to outweigh the costs of enacting the mitigating measure. In practice, however, it is difficult for users to have a comprehensive view of the situation and thus make informed decisions. Typically developers of secure systems best understand the nature of risks users will encounter and design responses that will mitigate those risks, but it is difficult for them to communicate this knowledge to users who are ultimately responsible for weighing risk severity and response cost trade-offs.

Consequently, the design of many security mechanisms seeks to simplify the threat-mitigation equation by avoiding calculations of risk impact and response cost, either through automating security measures or by pushing users to unilaterally enact protective measures regardless of context. This approach, however, is not without drawbacks. It discounts the impact of response costs on users' lives by presupposing that the execution of a protective behavior is always a favorable cost-benefit proposition. In reality, however, the "appetite and acceptability of a risk depends on [users'] priorities and values" [12]. Indeed, it has been argued that, "Security that routinely diverts the attention and disrupts the activities of users in pursuit of these goals is thus the antithesis of a user-centered approach" [20].

This approach and its drawbacks is evident in the current design of secure messaging applications. In a typical secure messaging application, an application server registers each user and stores their public key. When a user wishes to send a secure message to someone, the application transparently retrieves the public key of the recipient from the server and uses it to automatically encrypt messages. However, because the server could deceive the user, either willingly or because it has been coerced by a government or hacked by an attacker, communicating parties must verify one another's public keys in order to preserve the cryptographic guarantees offered by end-to-end-encryption. The method by which parties verify their public keys has been called the authentication ceremony, and typically involves scanning a contact's QR code or making a phone call to manually compare key fingerprints.


*USENIX Symposium on Usable Privacy and Security (SOUPS)* 2019.
August 11–13, 2019, Santa Clara, CA, USA.

The usability of the authentication ceremony in secure messaging applications has been studied in recent years, with the general conclusion that users are vulnerable to attacks, struggling to locate or perform the authentication ceremony without sufficient instruction [1, 21, 28]. The root cause of this difficulty is that the designers of these applications do not effectively communicate risks, responses, and costs to users. The automatic encryption "just works" when there is no attack, but the application does not give users enough help to judge risk and response trade-offs when an attack is possible. Prior work [29] applied opinionated design to the Signal authentication ceremony and showed that they could significantly decrease the time to find and perform the authentication ceremony, with strong adherence gains. However, this work assumed that all users should perform the ceremony for every conversation, when many users may not want to incur this cost due to low perceived risk or high response cost.

In this study, we demonstrate an alternative design approach that emphasizes users' comprehension over compliance, with a goal of empowering users to make more informed decisions that align with their personal values. We employ a design philosophy that might be seen as partway between opinionated and non-opinionated design: our design pushes users to make decisions, but not any decision in particular.

To this end, we conduct a three-phase redesign of the warning notifications surrounding the authentication ceremony in the Signal secure messaging app. We use Signal because the Signal protocol has been the foundation upon which other secure messaging applications have built, and thus many secure messaging applications share its basic design features and have similar authentication ceremonies. Because Signal is open source, we can apply design changes and, if these changes are successful, influence applications based on Signal, such as WhatsApp and Facebook Messenger.

The authentication ceremony in Signal is a particularly good fit for applying a risk communication approach to design. First, the system has an explicit and timely heuristic for identifying shifts in risk levels: encryption key changes. Moreover, because changes in security state are contingent upon key changes, we need only communicate with users once a potential risk occurs. Furthermore, the available mitigating response to a key change is unambiguous: performing the authentication ceremony. Finally, the authentication ceremony is a mechanism where response cost factors heavily into the equation—users must be synchronously available to perform it—even though most key changes are due to reinstalling the application, not a man-in-the-middle attack.

Our redesign generally follows a standard user-centered design process, but with an explicit focus on enabling users to make more informed decisions. First, we measured the baseline effectiveness of Signal's man-in-the-middle warning notifications with a cognitive walkthrough and a lab-based user study. Next, we designed a set of candidate improvements and evaluated their effectiveness by having participants on Amazon's Mechanical Turk platform interact with and rate design mockups. Lastly, we implemented selected improvements into the Signal app and evaluated our redesign with a user study that repeated the conditions of the first study.

We make the following contributions:

- **Identify obstacles to user understanding of the authentication ceremony in Signal.** We performed a cognitive walkthrough of Signal's authentication ceremony and associated notifications, highlighting barriers to understanding its purpose and implications. We followed up on our findings with a user study exposing participants to a simulated attack scenario, which allowed us to evaluate the effectiveness of these warnings in practice.

- **Perform a comprehension-focused redesign of the authentication ceremony with an aim at empowering users to balance risk-response trade-offs in a manner concordant with their personal priorities.** Building on the findings of our cognitive walkthrough and user study, we redesigned the authentication ceremony and associated messaging with a focus on empowering users to make more informed decisions. Candidate designs were evaluated by users on Amazon Mechanical Turk with a final redesign evaluated in a user study. Our redesign results in higher rates of both comprehension and adherence as compared to Signal's default design.

- **Show that risk communication empowers users to decide that not enacting protective behaviors is the right choice for them.** We find evidence that making users aware of the presence of an active threat to their data privacy is insufficient to produce secure behaviors. Users instead weigh the perceived impact of negative outcomes against the cost of enacting the response. Because "worst-case harm and actual harm are not the same" [10], this balancing of trade-offs can weigh unfavorably against performing protective measures.

- **Show that users' strategies for mitigating perceived threats are dependent on their perception of the larger security ecosystem as a whole.** Despite our redesign prompting a greater share of users to perform the authentication ceremony, and producing greater understanding of the purpose thereof, participants' preferred strategies for mitigating the perceived interception risk did not change substantially. Instead, it is apparent that users have developed an array of protective behaviors they rely upon to ensure positive security and privacy outcomes that exist beyond the ecosystem of any given app or system.

**Artifacts:** A companion website at https://signal.internet.byu.edu provides study materials, source code, and anonymized data.

## 2 Related work

### 2.1 Protection motivation theory

We base our work on protection motivation theory (PMT), which tries to explain the cognitive process that humans use to change their behavior when faced with a threat [14, 19]. The theory posits that humans assess the likelihood and severity of a potential threat, appraise the efficacy and cost of a proposed action that can counter the threat, and consider their own efficacy in being able to carry out that action.

Recently, PMT has been applied to a variety of security behaviors. Much of the work in this area is limited to studying the *intention* of individuals to adopt security practices, such as the intention to install or update antivirus software, a firewall, or use strong passwords [13, 32]. However, psychological research has demonstrated there is a gap between intention and behavior [22, 23], similar to the gap reported between self-reported security behaviors and practice [30]. A few studies have used objective measures of security behavior to study connections to PMT, such as compliance with corporate security policies [32], adoption of home wireless security [31], and secure navigation of an e-commerce website [27].

### 2.2 Risk communication

We are interested in studying how application design can be modified to help users assess risk and thus make more informed choices. We thus draw upon the wide variety of work in usable security that has focused on the design of warnings given to users.

Microsoft developed the NEAT guidelines for security warnings [18], emphasizing that warnings should only be used when absolutely *necessary*, should *explain* the decision the user needs to make, should be *actionable*, and should be *tested* before being deployed. Browser security warnings, in particular, have had a long history of lessons learned, including eliminating warnings in benign situations [26], removing confusing terms [4], and following the NEAT guidelines [8]. Phishing warnings are recommended to interrupt the primary task and provide clear choices [6]. Other work has recommended that software present security behaviors as a gain and use a positive affect to avoid undue anxiety [9].

We also draw upon risk communication, a discipline focused on meeting the need of governments to communicate with citizens regarding public health and safety concerns [5]. Nurse et al. provide a summary of how risk communication can be applied to online security risks [16]. Their recommendations include focusing on reducing the cognitive effort by individuals, presenting clear and consistent directions for action, and presenting messages as close as possible to the risk situation or attack. One noteworthy effort used a risk communication framework to redesign warnings for firewall software [17]. Their results show that the warnings improved comprehension and better communicated risk and consequences. However, the focus of this study, as with many others, was on greater compliance with recommended safe behaviors.

In contrast, we feel that risk communication provides a greater benefit in usable security when it enables users to make rational decisions based on their values, as opposed to compliance with a prescriptive behavior that experts believe is correct. For example, Herley has emphasized the rationality of users' rejection of security advice, by explaining that users understand risks better than security experts, that worst-case harm is not the same as actual harm, and that user effort is not free [10]. Sasse has likewise warned against scaring or bullying people into doing the "right" thing [20]. Indeed, recent work on what motivates users to follow (or not follow) computer security advice indicates that differences in behavior stem from differences in perceptions of risk, benefits, and costs [7].

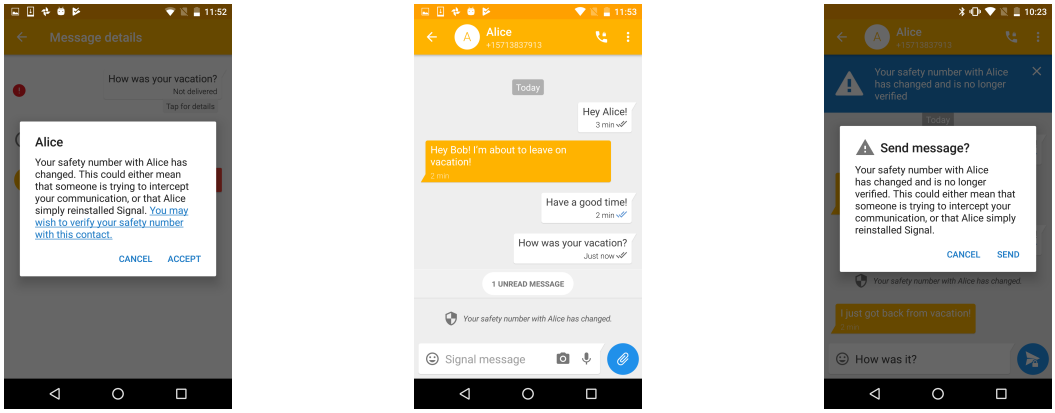
As stated by the National Academies, *"citizens are well informed with regard to personal choices if they have enough understanding to identify those courses of action in their personal lives that provide the greatest protection for what they value at the least cost in terms of those values"* [5]. Success is measured in terms of the information available to decision makers, and need not result in consensus or uniform behavior due to differences in what individuals value or perceive in terms of risks or costs of action.

## 3 Evaluating warnings in Signal

Signal uses the phrase *safety number* to describe a numeric representation of the key fingerprints for each participant in a conversation, warning users when this safety number changes. A safety number change occurs either when someone reinstalls the app (which generates new keys), or if a man-in-the-middle attack is conducted, with an attacker substituting their own key for an existing one. The authentication ceremony in Signal is referred to as *verifying* safety numbers; matching safety numbers rules out an attack. To evaluate the effectiveness of notifications that Signal currently uses we conducted both a cognitive walkthrough and a lab user study.

### 3.1 Cognitive walkthrough

We performed a cognitive walkthrough of the notifications presented to users when a key change occurs and the authentication ceremony. The walkthrough was conducted by four of the authors, with a range of experience—a professor and a graduate student with substantial prior HCI and Signal research and two undergraduate students with no prior experience with HCI or with Signal. Our walkthrough consisted of exposing the user to every possible scenario leading to a safety number change, documenting all notifications and messages that are presented to the user and mapping the flow of decisions the user can make at each point. In addition, we




(a) Message not delivered dialog (b) Shield message (c) Message blocked dialog

Figure 1: Signal notifications when safety numbers differ, depending on the internal state of the application.

analyzed Signal's code base to establish how internal state accompanied each warning notification and the effects of user actions on these states.

Our cognitive walkthrough revealed that, depending on the internal state of the system prior to a key change, Signal will react in one of three different ways to a key change event, as depicted in Figure 4 in Appendix A:

- *Message not delivered (top path in Figure 4):* This path is activated when the user has not previously verified safety numbers, is still on the conversation screen, and attempts to send a message. Sent messages will show up in the conversation log, accompanied by a notification informing the user that they were "not delivered" and that they may tap for more details. Doing so brings up another screen which clarifies that there is a "new safety number" alongside a "view" button. Tapping the button generates a dialog (Figure 1a) with a succinct message about safety number changes and several options for proceeding, including one that leads to the authentication ceremony screen and one that clears the warning state.

- *Message delivered (bottom path in Figure 4):* This path is activated when the user has not previously verified safety numbers and has either left the conversation screen or received a message. Signal will insert a notification into the conversation log informing the user of a safety number change, using a shield icon to mark the notification (Figure 1b). Tapping this dialog will take the user to the authentication ceremony screen. The shield and message appear in all three flows, but this is the only notification given to users in this flow; no other changes occur.

- *Message blocked (middle path in Figure 4):* This path is activated when the user has previously verified safety numbers and has either left the conversation screen or received a message. This scenario places a blue banner at the top of the conversation log, warning users that their "safety number has changed and is no longer verified". Tapping this banner takes users to the authentication ceremony. If the user attempts to send a message while in this state, Signal will prevent the message from being sent, and a dialog will be shown (Figure 1c). This dialog informs users that the safety number has changed and asks whether they wish to send the message or not. The user has three ways to clear the warning state in this scenario. They may select the "send" option at the dialog, mark the contact as verified on the authentication ceremony screen, or tap the "x" on the blue banner.

Our cognitive walkthrough identified numerous issues that may be confusing and that contradict recommendations on effective warning design:

- *Unclear risk communication.* It may not be clear to users what the term "safety number" means, nor what it means that these have changed.

- *Inconsistency of choice across dialogs.* Although the message-not-delivered and message-blocked flows show dialogs that convey nearly identical messaging, they present users with different choices for interaction (Figures 1a and 1c respectively).

- *The consequences of user actions are not clear beforehand.* For example, in the message-not-delivered flow, it is likely for the user to send multiple messages that are blocked from delivery before noticing and attempting to resolve the error. If the user selects "Accept" at the ensuing dialog, this will automatically re-send *all* failed messages; not just the one selected for inspection. Conceivably, should one or more of those failed messages contain sensitive information, this might be undesirable behavior.

- *The implications of success or failure of the authentication are unclear.* In the event of a failed safety number

match—the identification of which is the entire reason for the authentication ceremony—no recommendations for subsequent action are made to the user.

- *Does not communicate response cost.* The costs and requirements for performing the authentication ceremony are not made clear before users are brought to the authentication ceremony screen.

## 3.2 User study #1: Methodology

The following study, and all others in this work, were approved by our Institutional Review Board.

We designed a between-subjects user study to evaluate the effectiveness of each of these three notification flows at informing users of the potential risks they face and the responses available to them when exposed to a man-in-the-middle attack scenario. To control environmental conditions all participants used a Huawei Mate SE Android phone that we supplied.

For each of the three notification flows we discovered in our cognitive walkthrough, 15 pairs of participants (for a total of 45 pairs) conducted two simple conversation tasks. A simulated man-in-the-middle attack was triggered between the first and second tasks, causing the corresponding warning notifications to appear for each participant at the start of their second task. We simulated the attack by modifying the Signal source code to contact a server we operate and then change the encryption keys on demand. Participant reactions were recorded with video and a post-task questionnaire.

Our choice of tasks differs from previous work that asked participants to transmit sensitive information. Instead, we had participants communicate non-sensitive information, because this has the potential to reveal more diverse behaviors when faced with a risk of interception. For example, some users may be unconcerned by interception or unwilling to incur the cost of conducting the authentication ceremony if they perceive a conversation with non-sensitive information to be low risk. Others, on the other hand, may still find a potential attack to be unsettling and thus assess the risk to be more severe and/or the cost to be more worthwhile. A scenario with sensitive information could interfere with this dynamic.

We performed the studies for each treatment type—each notification flow—in succession, such that the first 15 pairs all experienced the message-not-delivered flow, the next 15 pairs saw only the message-delivered flow, and the final 15 pairs were exposed to the message-blocked flow.

### 3.2.1 Recruitment and Demographics

We recruited participants by posting flyers in buildings on our university campus. The flyer instructed participants to bring a partner to the study. Participants were each compensated $15, for a total of $30 per pair. Studies lasted approximately 40 minutes.

Our sample population skewed young, with 92.2% (n=83) of our participants aged between 18-24. Our population also skewed female (61.1%, n=55). A skills-based, self-reported assessment of technical familiarity revealed a normal distribution with most participants familiar with using technology.

### 3.2.2 Study design

When participants arrived, they were randomly assigned to an A or B roleplay condition (with a coin flip). Participants were then escorted to separate rooms, where they were presented with a packet of instructions, with one page per task.

Participants were first directed to register the Signal app pre-installed on the phones, granting all permissions the app sought in the process. Once both participants had finished registration, they were directed to begin their first task: to coordinate a lunch appointment using Signal. This task was designed to familiarize our participants with the operation of Signal. Exchanging messages is also necessary for Signal to establish safety numbers that could then be changed as part of the man-in-the-middle-scenario.

Next, participant B's roleplay informed them that participant A had gone to Hawaii on vacation, and to hand their phone to their study coordinator to simulate this communication disconnect. Participant A's roleplay provided similar information, including the instruction to hand their phone to their study coordinator, but additionally provided a half-page description of their "trip".

Study coordinators took this opportunity to manipulate Signal into the conditions necessary for the associated treatment as well as triggering the simulated man-in-the-middle attack. Finally, phones were handed back to participants, and they were instructed to continue on to their final task.

Finally, participants were instructed to discuss and share photos of participant A's trip to Hawaii, which had been preloaded onto participant A's phone. With the simulated attack active, participants were now exposed to the warning notifications corresponding to their treatment group. These final instructions explicitly stated that participants were finished with this task whenever *they* believed they were, to avoid biasing participants toward any particular action in the event of a failed authentication ceremony.

Once both participants declared the task complete, they were given the post-task questionnaire. This questionnaire asked them if, within the context of their roleplay, they had perceived a risk to their privacy. They were then asked how they might mitigate this risk, and to describe how effective they believe their strategy would be. Finally, participants were shown each of the warning notification elements in turn, and asked: (1) whether or not they had seen them, (2) what message they believed the notification was attempting to convey, and (3) what effects they believed the associated interactive elements would produce.

Upon completion of the questionnaire, participants were read a short debrief, informing them that the attack had only been simulated, that Signal employs multiple features intended to both prevent and identify interception, and that no such attacks have ever been reported in the wild.

#### 3.2.3 Data analysis

All open-ended questionnaire responses were coded by two of the authors in joint coding sessions using a conventional content analysis approach [11].

### 3.3 User study #1: results

#### 3.3.1 Risk perception and mitigation

Roughly half of groups 1 and 3, the treatment groups whose messages either failed to send or were blocked, perceived a risk during the study scenario (13/30 and 16/30 participants respectively). In stark contrast, however, only a small fraction of the participants in group 2 (4/30), whose workflow was not interrupted, felt that they had encountered a risk. In explaining the nature and properties of the risk they perceived, participant responses generally fell in one of three categories: (1) a security risk of an unknown nature, (2) a risk of interception, or (3) a risk of an insecure communication channel. Perceptions of how to mitigate such a risk generally fell under one of three categories: self-filtering (avoiding communicating sensitive information), use of an alternative communication channel such as another app, and verifying a contact.

#### 3.3.2 Shield message

The shield message in the conversation log, "Your safety number with <contact> has changed", confused a number of participants. While many participants correctly associated this message with a change in security status, a number interpreted it to mean precisely the opposite of its actual meaning—that it conveyed *improved* security levels. As one participant explained following our post-study debrief, *"I thought that it was improving security—that every once in a while, you change the safety number so it refreshes and makes it harder for people to hack into. So, I was like, 'Oh, it's doing its job.' Apparently, it wasn't!"*

Next, as our cognitive walkthrough predicted, participants were confused by what, precisely, it was that had changed, offering numerous different explanations. Examples include: phone number, connection, safety number, safety code, "something technical", settings, security code, and verification code. As one participant remarked, *"Some sort of safety code changed. Or his actual phone number, I was a little confused."*

Participants acted on this message all cited the importance of ensuring privacy/security outcomes. Those who did not act on it did so because: (1) they did not see it as an actionable message, (2) they explicitly expressed having been habituated against such notifications, (3) the information they were communicating was seen as non-sensitive, or (4) they perceived it to be a part of the study task.

Notably, perceptions of the non-sensitivity of the conversation were critical in putting participants at ease even if they had found the notification alarming, as exemplified by one participant response: *"I felt that it was important because of the nature of the app and whenever a safety anything is changed that usually is noteworthy. I would have put that it was extremely important if I had felt like there was an actual risk of someone actually trying to read our conversation."*

#### 3.3.3 Message-not-delivered dialog

Only participants in treatment group 1 were exposed to the message-not-delivered dialog. Participants were asked to describe what they believed would happen if they were to tap the three interactive elements in this dialog: the "Accept" and "Cancel" buttons and the link embedded in the text.

Participants generally understood that "Cancel" would leave the system state unchanged. Similarly, most participants understood that "Accept" would unblock their messages and allow them to communicate once more. Perception of the link, however was more confused. 9 of the 14 participants who responded to this question responded that they had believed it would have taken them to a screen explaining more about the situation. This is in contrast to what it really does, which is to redirect users to the authentication ceremony, as noted by one participant who expected it to lead *"to an 'About' or 'Info' page, but it ended up taking me to the verification."*

#### 3.3.4 Blue banner & message-blocked dialog

Understanding of the options presented by the message-blocked dialog—"Send" and "Cancel"—were high. However, unlike the message-not-delivered dialog, the message-blocked dialog does not present a method to reach the authentication ceremony—instead accessible via the blue banner.

Understanding of the blue banner was mixed among those participants of group 3 who reported having seen it. Only roughly half understood that it was a privacy-related warning. Others were either entirely at a loss to explain its purpose or believed that it was a system error notification. Those who were confused by its meaning or believed it to be a system error did not feel it warranted action. Of the five participants who correctly interpreted the blue banner as a warning, two did not feel they were at risk, and thus did not feel like action was warranted.

#### 3.3.5 Authentication ceremony

Participants who reported having seen the authentication ceremony screen were asked about the significance of verifying safety numbers (and whether or not they matched) as well as about the verification toggle. Participants may have seen,

and even interacted with, the authentication ceremony screen without necessarily having performed the authentication ceremony. In total, 5 pairs of participants conducted the authentication ceremony while 27 participants reported having seen the screen.

As predicted in our cognitive walkthrough, participants were confused about what a safety number was or why it had changed. For instance, one participant explained that *"I honestly wasn't sure what it meant. I didn't know that I had a safety number with them in the first place so I was unaware that it could change."* We also noted occasions where participants entered the authentication ceremony screen only to back out without completing it. This may be due to poor communication regarding response cost—both conversation partners should either be in the same physical location to execute the QR-code ceremony or be willing to verify safety numbers over another medium (such as a phone call).

Also as predicted, the verification toggle confused participants. Of the 11 participants who reported having flipped the toggle, not one participant correctly intuited its use. 7 of these 11 toggled it purely as an exploratory action, unaware that doing so would inadvertently and incorrectly clear the warning state.

When asked to characterize the purpose of the authentication ceremony, participants did generally associate it with verification, although their model for *what* it verifies was often incorrect. Table 2 shows a qualitative analysis of participant responses when asked the purpose of the ceremony, and the meaning of a matching or non-matching result, with responses coded and then categorized as correct, partially correct, or incorrect. Only a few participants understood that the purpose of the authentication ceremony is to verify the confidentiality of the conversation. Instead, a number of participants mistakenly believed that it was about verifying the identity of the individual, i.e., that *"it makes sure the other person is who you think they are"*, as one participant explained. This threat model does not account for a different type of attacker the authentication ceremony is intended to detect: a passive man-in-the-middle who simply decrypts and forwards messages without interfering in the conversation.

These misconceptions naturally carried forward into responses about the significance of matching and non-matching safety numbers. Perceptions of non-matching safety numbers correctly assessed this result as indicative of interception occurring, but again, participants often believed that this meant that they had detected an impersonator, as with one participant who remarked that, *"Someone using another phone could be posing as my brother, I guess."* Participants did almost unilaterally understand that matching safety numbers were indicative of a positive security/privacy outcome, although several participants misinterpreted the role of the authentication ceremony as a mechanism that would actively *prevent* interception, as opposed to detecting it.

# 4 Developing improvements

Based on the results of our cognitive walkthrough and subsequent user study, we concluded that there were three main areas for improvement worthy of focus: (1) the need for an accessible, persistent visual indicator for verification state, (2) the messaging used in warning notifications and dialogs, and (3) the notification flow and all associated UI elements.

## 4.1 Visual indicator

Visual indicators, or icons, are important both as an accessible measure for communicating security state to users with a single glance as well as for enhancing the consistency of warning notifications. While the authentication ceremony screen in the original version of Signal does have a (somewhat hidden) lasting representation of verification state, the verified toggle switch, we believe that this indicator is inadequate because it represents only two states (verified and unverified) and because it confused users in our lab study who believed that toggling the switch would verify their partner.

We decided to create a set of icons that would properly reflect all three verification states: (1) the default, assumed-safe state of the conversation prior to a safety number change, (2) a verified state that reflects matching key fingerprints, and (3) an unsafe state that reflects having found non-matching fingerprints in the authentication ceremony. Ideally, the icon for the default state could have a small modification to represent the other two states. By adding this visual indicator onto the action bar, it becomes both an accessible indicator of state as well as a shortcut to the authentication ceremony.

We began by designing a neutral icon to represent the default state. Our goal was to select an icon that would be intuitively associated with privacy, and that would not evoke unwarranted feelings of concern, since this state does *not* signal a cause for concern. We selected a blank shield icon for this purpose. We then created variants of this icon, as shown in Table 1 to represent the success and failure states post-authentication ceremony.

We evaluated our designs on Amazon's Mechanical Turk platform, with each icon being shown to at least 50 participants. Each icon was shown occupying a position on the action bar in a screenshot of Signal's interface, next to the call button. For positive-valenced icons we asked participants to rate how strongly they associated the icon with privacy on a scale from 1-10. For negative-valenced icons we asked participants to rate how worried they would feel if they saw the associated icon. We asked both questions for the blank shield icon.

As shown in Table 1, the blank shield has a moderate association with privacy and a low association with worry, making it a good fit for a default icon. We discounted any icons using a lock because it is used elsewhere in the app to represent encryption, and we wished to avoid conflating meanings. We

Table 1: Comparison of icons a 10-point Likert scale

| Icon | Mean | Std. Dev. | Count |
|---|---|---|---|
| *Positive – association with privacy* | | | |
|  | 6.50 | 2.21 | 74 |
|  | 6.54 | 2.82 | 79 |
|  | 5.74 | 2.56 | 78 |
|  | 7.52 | 2.43 | 83 |
| *Negative – association with worry* | | | |
|  | 4.08 | 2.36 | 59 |
|  | 4.95 | 2.36 | 55 |
|  | 5.11 | 2.64 | 57 |
|  | 4.17 | 2.51 | 65 |
|  | 4.95 | 2.49 | 60 |
|  | 4.52 | 2.53 | 52 |
|  | 5.56 | 2.53 | 61 |
|  | 5.05 | 2.51 | 62 |

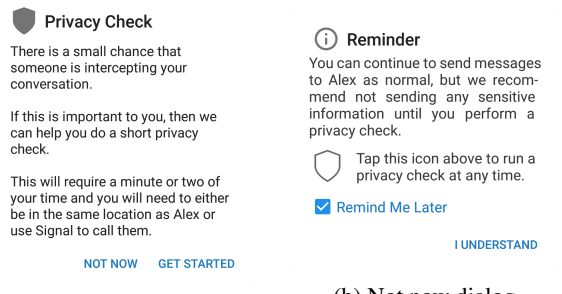
chose the shield with a checkmark enclosed by a circle for the positive icon because of the remaining choices it had the strongest association with privacy. Surprisingly, no negative icon evoked strongly negative associations. We chose the shield with an exclamation mark because it had the strongest negative associations, and if the privacy check fails we do want users to be alarmed.

Appendix C shows how these indicators are used in our design.

## 4.2 Notification and dialogs

We revised notifications and dialogs concerning safety numbers throughout Signal by following recommendations for warning design and risk communication. The principles we followed are (a) interrupt the primary task, (b) present messages close to the risk situation, (c) reduce cognitive effort (d) use a positive affect, (e), explain the decision the user needs to make, and (f) present clear and consistent directions for action. In particular, we designed the following changes, with screenshots shown in Appendix B:

- *Positive framing for the authentication ceremony.* We framed the authentication ceremony as a "privacy check", which emphasizes the *role* it plays rather than the primitives or actions involved, which will be unfamiliar to users. Notifications of changed safety numbers (what we refer to as the "shield message") instead report that Signal recommends a privacy check, turning what was sometimes seen as a routine system notification into an explicitly actionable recommendation. We also frame the consequences of performing the privacy check in a positive manner such that both positive and negative results present benefits for the user: a positive match guarantees conversation privacy and a failed match reveals ongoing message interception. In this way, even if fingerprints do end up matching—by far the most common case—users need not feel that it was a waste of their time to engage in the verification process.




(a) Privacy check dialog

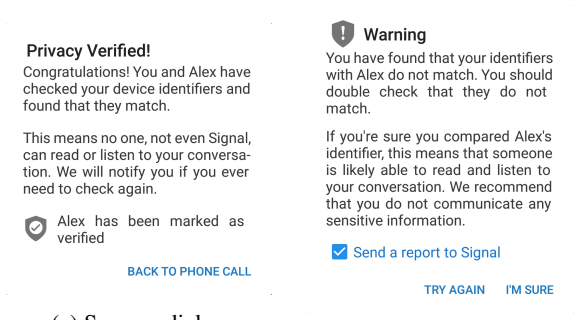
(b) Not now dialog

Figure 2: New notification dialogs, framing the authentication ceremony as a privacy check and using risk communication principles.

- *Communicating response cost and providing users with alternatives to the privacy check.* Our dialogs inform participants up-front what executing the privacy check will require (Figure 2a), with a "Not now" option that generates a reminder dialog for participants who are uninterested (Figure 2b). The reminder dialog includes a recommendation to not communicate sensitive information until the privacy check is first completed, with an option be reminded at a later time and a description of how to access this functionality at any time. Thus, participants' options are framed as clear choices with defined costs.

- *Safety number labeling and interaction changed.* To promote better understanding of the safety numbers and their role, we divided the safety numbers into their constituent halves, relabeled them as *device identifiers*, and explained that they are used in encrypting the conversation.[1] Prior work indicated users dislike how long the safety number is [28]. Thus, we also rearranged groupings from 5 digits in a set to 3. This aligns more naturally with the standard process for grouping numbers, where numbers larger than 999 are grouped into sets of three known as periods. This does not reduce the actual count of numbers, but does reduce cognitive load.

[1] This is not technically accurate, as they are key fingerprints and not keys themselves, but our goal is to have participants associate the comparison task with the preservation of a secure conversation and not to overwhelm them with details of the encryption process.




(a) Success dialog

(b) Failure dialog

Figure 3: New authentication ceremony success and failure dialogs.

We provide two options for performing the authentication ceremony, an in-person QR-code scan and a phone call comparison, as recommended in [29]. In the phone call version, we removed the confusing toggle element, which we replaced with two buttons explicitly labeled "Match" and "No match".

- *Addition of success and failure messaging when the privacy check is completed.* We added dialogs after the privacy check that inform users of the implications of success and failure (Figures 3a and 3b respectively). We also designed a dialog shown before the authentication ceremony. If the privacy check has already been completed, it will show the current state and its implications for the user; if it has not, then it will instead explain what the privacy check entails and provide access to our authentication ceremony options.

- *Options for interaction inform users of the choice they're making.* To promote user autonomy and informed decisions, we carefully selected labels for our dialog buttons that describe the consequences of that choice and imply active decision-making on the part of the user, such as "Not now" and "Get started" on the privacy check dialog, as opposed to the more traditional "Okay" and "Cancel".

### 4.3 Notification flow

As described earlier, based on the system state prior to a key change, Signal diverges into one of three different notification flows. In order to provide a consistent user experience, we decided to instead use a single, unified flow every time a key change occurs. We eliminated the non-interrupting flow from consideration because in our first study it was ineffective at promoting either adherence or comprehension. This left us with the two interrupting flows, the message-not-delivered and message-blocked flows, which produced similar comprehension levels in our user study. We hypothesized there might be a difference between these because the timing of interruptions can have an impact on decision-making [2, 3, 15, 25]. We further identified two additional UI elements that might contribute to our aims of increased comprehension: (1) an introduction screen showing the privacy check icons after registration and (2) the blue banner element that accompanies the message-blocked flow in the original Signal.

To evaluate the relative effectiveness of these elements we designed a website containing a simulated Signal experience using mockups of our candidate flows and had users of Amazon's Mechanical Turk platform interact with the simulation using a between-subjects comparison. Participants were, as in our user study, presented with two simple communication tasks involving non-sensitive information and a man-in-the-middle occurring between the first and second task. Unlike in our user study, users selected from a set of predefined messages, although they otherwise interacted with the interface normally, and their interactions were recorded in a database. For example, participants that wished to proceed with the privacy check were shown the results of the check as if they performed the authentication ceremony and asked to choose a response from the resulting dialog. After they were done, participants were given a tailored questionnaire which asked their perception of the notifications they had seen as well as why they had chosen the options they did. A total of 223 participants interacted with mockups and explained their actions via a post-task questionnaire.

We separated the elements to be evaluated into three rounds. The first round compared our delivery mechanisms: the message-not-delivered and message-blocked flows. The winner of the first round was then evaluated against a version that also included the blue banner element. Finally, the winner of the second round was then evaluated against a version that added an introductory screen.

To test for the difference between the message-not-delivered and the message-blocked flows we measured how many participants chose to start the authentication ceremony. We observed no significant difference (35/50 vs 31/50). We opted to use the message-blocked control flow because the message-not-delivered flow complicates the user's task when they must resolve multiple failed messages. To test whether the blue banner message had an improvement we again measured how many participants chose to start the authentication ceremony and observed no difference (31/50).

To test whether the introductory screen had a difference we qualitatively measured comprehension. To do this we used participant responses to a question asking them what the privacy check notification meant. Several authors coded each response and then determined whether the participant understood that this notification meant interception of their conversation could be happening and found a slight improvement with the introductory screen (30/50 vs 23/50). However, we chose to leave the introductory screen out of our final design because the effect was not large and could have been exaggerated due to the short-term nature of the simulation.

Qualitative analysis of participant responses regarding their decision to perform (or not perform) the privacy check showed

participants weighed risks with response costs and made reasoned choices. Roughly 60% of all groups opted to perform the privacy check, with the remainder choosing the other option, "not now". Participants who opted to perform the privacy check typically stated having done so out of a desire to verify the existence of the risk, because they believed it better to be safe than sorry, or out of curiosity. Participants who chose "not now" had either determined the risk to be of minimal severity or because they felt executing the privacy check would be inconvenient. Those who felt it would be inconvenient described it as such either because the current timing was seen as inappropriate or because of the synchronization cost (needing both members of the pair to execute the privacy check at the same time).

# 5 Evaluating the effectiveness of our redesign

We conducted a lab study to evaluate the effectiveness of our changes. Appendix B shows the control flow we used and screenshots of the new notifications and UI elements, and Appendix C shows the new indicators. We maintained the same study design used in our first lab study, with some minor modifications. Since our redesign has just one control path, we ran this study with just one treatment group of 15 pairs. Because we included additional screens that have no analogous equivalent in the original version of Signal, the post-task questionnaire in this study is not fully comparable with that from our first.

## 5.1 Results

### 5.1.1 Risk perception

Two-thirds (20/30) of our final user study participants reported having perceived a threat within the context of their roleplay. Qualitative responses indicated participants largely correctly perceived an interception risk, while a handful, interestingly, believed that Signal was itself the risk; this view seemed to be fueled by the number of permissions that Signal asks for in short succession. Participant notions of how they might mitigate perceived risks virtually mirrored those from the first user study—self-filtering, using alternate communication channels, and verifying contacts—along with restricting app permissions. Notably, only a small fraction of open responses (2/20) mentioned the privacy check as their mitigating strategy of choice despite, as we describe shortly, improved adherence and comprehension rates.

### 5.1.2 "Signal recommends a privacy check"

Qualitative responses indicate nearly all participants associated this notification with security, although as with our first study, there were a few who misinterpreted it as an *increase* in security. Due to our removal of Signal's original messaging regarding a change, participants of our final user study were not confused about what had changed as the first groups had been.

Those who felt it important to act upon this message generally explained that they felt ensuring privacy outcomes to be important, as with one participant who explained, *"I hear a lot about data breaches and such, so seeing that the app was giving a warning notification showed to me that it was something important that I should act on."* Importantly, those who did not feel that the notification was cause for concern typically felt that way because the information they were communicating was perceived as non-sensitive in nature.

### 5.1.3 Privacy check dialog

Qualitative responses indicate participants generally understood the dialog was informing them of a potential threat, although perceptions of the nature of that threat and of the likelihood of that threat were more varied. For example, while most participants correctly perceived that the dialog informs them only of a "potential" threat, a couple participants misinterpreted this notification as informing them of a confirmed threat, as with one participant who believed that *"someone was hacking my account"*.

Qualitative analysis of participant responses regarding their decision to perform (or not perform) the privacy check showed participants weighed risks with response costs and made reasoned choices. These results roughly match those of the Mechanical Turk participants who evaluated our candidate designs. Participants who felt performing the privacy check was important reported that this stemmed out of a desire to confirm the validity of the reported risk or because they believed it better to be safe than sorry. Those who did not feel the privacy check worth doing, on the other hand, had either deemed the risk minimal or decided that conducting the privacy check was too inconvenient.

### 5.1.4 Privacy check

As with the authentication ceremony in the first user study, participants in our final user study may have seen and interacted with the privacy check screen without having conducted the privacy check itself. 17 of our 30 participants reported having seen the privacy check screen, with 3 participants unsure. 6 participant pairs fully performed the privacy check, while 3 participant pairs partially performed the check (one participant in each pair incorrectly informed their partner that they had already matched the identifiers and that they thus did not need to complete the full process). This is in contrast with the 5 pairs (out of 45) who performed the authentication ceremony in our first study.

These three "successful" misunderstandings had the same root cause—our design was not robust against false positives. Our design pops up an informative success dialog when a user

Table 2: Comparison of participant understanding of the authentication ceremony using Signal and our redesign.

| Auth. cerem. | Signal | | Redesign | |
|---|---|---|---|---|
| Correct | Verifies security (conversation) | 2 | Verifies security (conversation) | 7 |
| | Verifies device | 2 | Verifies device | 1 |
| | | **4 [16%]** | | **8 [50%]** |
| Partially correct | Verifies person (not impersonator) | 8 | Verifies security (connection) | 3 |
| | Verifies security (connection) | 1 | Verifies person (not impersonator) | 2 |
| | Prevents interception (connection) | 1 | Prevents interception (conversation) | 2 |
| | Improved security (conversation) | 1 | Verifies security | 1 |
| | | **11 [44%]** | | **8 [50%]** |
| Incorrect | Don't know | 5 | | |
| | Verifies phone number | 2 | | |
| | Verifies security (phone) | 1 | | |
| | Prevents robocalls | 1 | | |
| | Makes the contact trusted | 1 | | |
| | | **10 [40%]** | | **0 [0%]** |

| Matching | Signal | | Redesign | |
|---|---|---|---|---|
| Correct | Interception not possible | 1 | Interception not possible | 6 |
| | Verifies device | 2 | Verifies device | 2 |
| | Verifies security (conversation) | 4 | Verifies security (conversation) | 1 |
| | | **7 [26.9%]** | | **9 [56.3%]** |
| Partially correct | Verifies person (not impersonator) | 7 | Verifies person (not impersonator) | 3 |
| | Improved security | 3 | | |
| | Prevents interception | 2 | | |
| | Prevents interception (conversation) | 2 | | |
| | Prevents interception (connection) | 1 | | |
| | Verifies security (connection) | 2 | | |
| | | **17 [65.4%]** | | **3 [18.8%]** |
| Incorrect | Don't know | 1 | Don't know | 3 |
| | Confusion | 1 | Confusion | 1 |
| | | **2 [7.7%]** | | **4 [25%]** |

| Non-matching | Signal | | Redesign | |
|---|---|---|---|---|
| Correct | Interception occurring (MITM) | 1 | Interception occurring (MITM) | 4 |
| | Interception occurring | 4 | Interception occurring | 2 |
| | Conversation not secure | 2 | Conversation not secure | 2 |
| | | | Device is impersonator | 1 |
| | | **7 [28%]** | | **9 [56.3%]** |
| Partially correct | Interception occurring (contact is impersonator) | 8 | Interception occurring (connection not secure) | 1 |
| | Connection not secure | 3 | Interception occurring (contact is impersonator) | 1 |
| | | | Connection not secure | 2 |
| | | **11 [44%]** | | **4 [25%]** |
| Incorrect | Don't know | 3 | Don't know | 2 |
| | App is not secure | 2 | Conversation is secure | 1 |
| | Robocalls | 1 | | |
| | Technical issues | 1 | | |
| | | **7 [28%]** | | **3 [18.8%]** |

taps the "Match" button. Unfortunately, this confused these participants who had mistakenly tapped the "Match" button. More specifically, one participant assumed that the "Match" button would activate an automated mechanism that would perform the verification for them. When the success dialog popped up in response, this participant assumed that the result had been in response to this "automated process". The other mistaken participants accidentally tapped the "Match" button and were similarly misled by the resulting success dialog.

Table 2 shows a qualitative analysis of participant responses when asked the purpose of the privacy check, and the meaning of a matching or non-matching result, with responses coded and then categorized as correct, partially correct, and incorrect. This table reveals that comprehension of the purpose of the authentication ceremony and of the significance of matching and non-matching numbers visibly improved with our redesign. While far from perfect, these results are promising given the context: a non-sensitive task scenario, no accompanying instruction or tutorials, and no incentive. Risk communication was limited to the messages contained within the application.

For all categories and for both user studies, partially correct responses center on the same few misconceptions: believing the verification process itself to be an active prevention mechanism, believing the "connection" and not the conversation to be the entity to be secured, and believing that the verification process verified the contact's identity, and not their device.

## 6 Discussion

### 6.1 Risk communication gives users the ability to make personal trade-offs between perceived risk and response cost.

Simply knowing that a negative outcome is likely to happen is not a sufficient reason to take action to prevent it: it must also be negative *enough*. As the participant quoted in the title of this work so eloquently stated, sometimes *"something isn't secure, but I'm not sure how that translates into a problem."* Indeed, this view was shared by a number of participants of our studies. We observed numerous instances where participants did not believe that conducting the authentication ceremony was a worthy use of their time, whether because they perceived their communications as non-sensitive and thus unworthy of protecting, or because they felt that performing the authentication ceremony would be too inconvenient. One shared response captures both these sentiments perfectly, *"If it was easy enough I would be happy to secure my conversation, but at the same time, how necessary is it?"*

While lowering response cost seems a natural way forward, particularly with automation, the deeply personal way in which calculations of risk function are made suggests obstacles ahead. Perceptions of risk severity in common scenarios will differ from person to person as a function of personal priorities and values. To wit, while many participants viewed communicating about our toy scenario as inherently non-sensitive, some participants were nevertheless uncomfortable at the realization that interception was "occurring". One such participant, commenting on the thought of an interceptor eavesdropping on their discussion of a fictitious Hawaii trip, remarked, *"Even though it's only about fish, that's not really cool with me."* We thus observe differences in risk assessment from different individuals although both the type of informa-

tion being communicated and the nature of the threat itself were identical in all cases.

For these reasons, it is our position that enabling users to truly make informed decisions requires properly communicating the nature and likelihood of the risk and the cost of recommended protective measures, and then giving them the freedom to determine that *not* actively protecting themselves is actually the decision most in line with their interests.

### 6.2 Users' strategies for coping with online threats extend beyond the ecosystem of your app.

Although our redesign evidenced both higher rates of participants conducting the authentication ceremony as well as comprehension, participants' responses of how they might mitigate the risks they had perceived did not change in any notable fashion. Despite both having been made aware of a protective measure (in the privacy check) and also having understood its purpose, participants ultimately did not find it a reliable measure for mitigating a perceived interception risk should they encounter a similar situation in the wild. Rather, participants mentioned self-filtering, restricting app permissions, and using alternative apps or channels of communication as key strategies for dealing with the interception threat introduced in the study.

This appears to be due to varying ideas about the source of the risk; in-app mitigating measures can only be depended on to do so much. Because the privacy check and associated messaging only informed users that conversation confidentiality had been violated, but not *how* that interception had been accomplished, users completed the process of threat assessment with personal interpretations of the origin of the interception risk. Relevantly, if the source of the risk is perceived to be outside the scope of the app—or even the app itself—it seems imprudent to rely on mitigating strategies that fall within the domain of the app.

System trust, perhaps unsurprisingly, appears to play a key role in this calculation. One participant response as to how they might better protect themselves is particularly ironic—they would forego use of Signal and *"use [a] secure messag[ing] app like Facebook Messenger"*. Facebook Messenger does not protect conversations with end-to-end encryption by default, unlike Signal. However, due to unfamiliarity with Signal, and trust in Facebook, this participant's preferred strategy would be to move from a secure messaging platform to a less secure one.

Future work could examine whether additional risk communication regarding the source of the threat could lead to improved understanding of the efficacy of the privacy check. System designers should also consider that users choose, to varying extents, appropriate responses to perceived threats, and that these include viable methods above and beyond what the system itself offers.

## 7 Limitations

Our cognitive walkthrough was thorough but limited to the expertise of the authors who participated in it. We ameliorate this by having a variety of backgrounds among those who participated, but a walkthrough performed by other experts or novices may find different issues with Signal's notifications. Our Mechanical Turk studies of icons and notification flows are limited to a simulated experience and thus may not match what users would feel or choose when interacting directly with the application. Our lab studies were limited to a young, college student population and may not generalize to a larger or more diverse population. Our Mechanical Turk results from the simulation provide some evidence that the results of the second lab study generalize to a larger, more diverse population. It would be helpful to study populations with different risk-cost trade-offs, such as immigrants or dissidents, and to ascertain that risk communication translates well to other cultures and languages. Our lab studies are also limited because users may act differently due to the Hawthorne effect [24]. Several participants made comments indicating this limitation was present, such as *"while it is very concerning to me that someone could be intercepting my conversation, I thought that it was just because it was in a study."* However, because the focus of our study was on comprehension as opposed to behavior, this effect may be less impactful in our study.

Aside from these more common issues, we also observed a bug in Signal's phone call functionality. The first time a Signal user makes an outgoing phone call, the caller is unable to hear audio although the recipient can hear clearly. Participants in our study simply redialed their partner when this occurred, typically chalking the issue up to a spotty wireless connection. This error, however, was present in the user studies evaluating both the original version of Signal and our redesign, so if this bug did have an effect, it likely existed in both cases, and thus is unlikely to have caused discrepancies in our results.

## 8 Conclusion

In this paper, we present the results of our experience redesigning the risk communication surrounding Signal's authentication ceremony for comprehension. Our three-part process reveals significant obstacles to understanding in Signal's current design, and demonstrates the effectiveness of applying risk communication principles to system design. Our user studies, which deliberately employ a non-sensitive communication task, provide evidence that users' decisions *not* to enact protective behaviors are actually conscious, informed decisions that are the product of balancing response cost and risk assessment. We further find that users rely on a host of protective behaviors that exist beyond the scope of any particular app or system, and that, consequently, responses to perceived threats may similarly exist outside of system designers' control.

## 9 Acknowledgments

The authors thank the reviewers and our shepherd for their helpful feedback. This material is based upon work supported by the National Science Foundation under Grants No. CNS-1528022 and CNS-1816929, and by the Department of Homeland Security (DHS) Science and Technology Directorate, Cyber Security Division (DHS S&T/CSD) via contract number HHSP233201600046C.

## References

[1] Ruba Abu-Salma, Kat Krol, Simon Parkin, Victoria Koh, Kevin Kwan, Jazib Mahboob, Zahra Traboulsi, and M Angela Sasse. The security blanket of the chat world: An analytic evaluation and a user study of Telegram. In *European Workshop on Usable Security (EuroUSEC 2017)*. Internet Society, 2017.

[2] Piotr D Adamczyk and Brian P Bailey. If not now, when?: the effects of interruption at different moments within task execution. In *SIGCHI Conference on Human Factors in Computing Systems (CHI 2004)*. ACM, 2004.

[3] Brian P Bailey and Joseph A Konstan. On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in Human Behavior*, 22(4):685–708, 2006.

[4] Robert Biddle, Paul C Van Oorschot, Andrew S Patrick, Jennifer Sobey, and Tara Whalen. Browser interfaces and extended validation SSL certificates: an empirical study. In *ACM Cloud Computing Security Workshop (CCSW 2009)*. ACM, 2009.

[5] National Research Council et al. *Improving risk communication*. National Academies, 1989.

[6] Serge Egelman, Lorrie Faith Cranor, and Jason Hong. You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In *SIGCHI Conference on Human Factors in Computing Systems (CHI 2008)*. ACM, 2008.

[7] Michael Fagan and Mohammad Maifi Hasan Khan. Why do they do what they do?: A study of what motivates users to (not) follow computer security advice. In *Symposium on Usable Privacy and Security (SOUPS 2016)*. USENIX, 2016.

[8] Adrienne Porter Felt, Alex Ainslie, Robert W Reeder, Sunny Consolvo, Somas Thyagaraja, Alan Bettes, Helen Harris, and Jeff Grimes. Improving ssl warnings: Comprehension and adherence. In *SIGCHI Conference on Human Factors in Computing Systems (CHI 2015)*. ACM, 2015.

[9] Vaibhav Garg and Jean Camp. Heuristics and biases: implications for security design. *IEEE Technology and Society Magazine*, 32(1):73–79, 2013.

[10] Cormac Herley. So long, and no thanks for the externalities: the rational rejection of security advice by users. In *New Security Paradigms Workshop (NSPW 2009)*. ACM, 2009.

[11] Hsiu-Fang Hsieh and Sarah E Shannon. Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9):1277–1288, 2005.

[12] Clinton M Jenkin. Risk perception and terrorism: Applying the psychometric paradigm. *Homeland Security Affairs*, 2(2), 2006.

[13] Doohwang Lee, Robert Larose, and Nora Rifon. Keeping our network safe: a model of online protection behaviour. *Behaviour & Information Technology*, 27(5):445–454, 2008.

[14] James E Maddux and Ronald W Rogers. Protection motivation and self-efficacy: A revised theory of fear appeals and attitude change. *Journal of Experimental Social Psychology*, 19(5):469–479, 1983.

[15] Daniel C McFarlane and Kara A Latorella. The scope and importance of human interruption in human-computer interaction design. *Human-Computer Interaction*, 17(1):1–61, 2002.

[16] Jason RC Nurse, Sadie Creese, Michael Goldsmith, and Koen Lamberts. Trustworthy and effective communication of cybersecurity risks: A review. In *Workshop on Socio-Technical Aspects in Security and Trust (STAST 2011)*. IEEE, 2011.

[17] Fahimeh Raja, Kirstie Hawkey, Steven Hsu, Kai-Le Clement Wang, and Konstantin Beznosov. A brick wall, a locked door, and a bandit: a physical security metaphor for firewall warnings. In *Symposium on Usable Privacy and Security (SOUPS 2011*. ACM, 2011.

[18] Rob Reeder, E Cram Kowalczyk, and Adam Shostack. Helping engineers design NEAT security warnings. In *Symposium On Usable Privacy and Security (SOUPS 2011)*. ACM, 2011.

[19] Ronald W Rogers. A protection motivation theory of fear appeals and attitude change. *The Journal of Psychology*, 91(1):93–114, 1975.

[20] Angela Sasse. Scaring and bullying people into security won't work. *IEEE Security & Privacy*, 13(3):80–83, 2015.

[21] Svenja Schröder, Markus Huber, David Wind, and Christoph Rottermanner. When SIGNAL hits the fan: On the usability and security of state-of-the-art secure mobile messaging. In *European Workshop on Usable Security (EuroUSEC 2016)*. IEEE, 2016.

[22] Paschal Sheeran. Intention—behavior relations: a conceptual and empirical review. *European Review of Social Psychology*, 12(1):1–36, 2002.

[23] Paschal Sheeran and Thomas L Webb. The intention–behavior gap. *Social and Personality Psychology Compass*, 10(9):503–518, 2016.

[24] Andreas Sotirakopoulos, Kirstie Hawkey, and Konstantin Beznosov. On the challenges in usable security lab studies: lessons learned from replicating a study on SSL warnings. In *Symposium on Usable Privacy and Security (SOUPS 2011)*. ACM, 2011.

[25] Cheri Speier, Joseph S Valacich, and Iris Vessey. The influence of task interruption on individual decision making: An information overload perspective. *Decision Sciences*, 30(2):337–360, 1999.

[26] Joshua Sunshine, Serge Egelman, Hazim Almuhimedi, Neha Atri, and Lorrie Faith Cranor. Crying wolf: An empirical study of SSL warning effectiveness. In *USENIX Security Symposium*. USENIX, 2009.

[27] René van Bavel, Nuria Rodríguez-Priego, José Vila, and Pam Briggs. Using protection motivation theory in the design of nudges to improve online security behavior. *International Journal of Human-Computer Studies*, 123:29–39, 2019.

[28] Elham Vaziripour, Justin Wu, Mark O'Neill, Ray Clinton, Jordan Whitehead, Scott Heidbrink, Kent Seamons, and Daniel Zappala. Is that you, Alice? a usability study of the authentication ceremony of secure messaging applications. In *Symposium on Usable Privacy and Security (SOUPS 2017)*. USENIX, 2017.

[29] Elham Vaziripour, Justin Wu, Mark O'Neill, Daniel Metro, Josh Cockrell, Timothy Moffett, Jordan Whitehead, Nick Bonner, Kent Seamons, and Daniel Zappala. Action needed! helping users find and complete the authentication ceremony in Signal. In *Symposium on Usable Privacy and Security (SOUPS 2018)*. USENIX, 2018.

[30] Rick Wash, Emilee Rader, and Chris Fennell. Can people self-report security accurately?: Agreement between self-report and behavioral measures. In *SIGCHI Conference on Human Factors in Computing Systems (CHI 2017)*. ACM, 2017.

[31] Irene Woon, Gek-Woo Tan, and R Low. A protection motivation theory approach to home wireless security. *International Conference on Information Systems (ICIS 2005)*, 2005.

[32] Michael Workman, William H Bommer, and Detmar Straub. Security lapses and the omission of information security measures: A threat control model and empirical test. *Computers in Human Behavior*, 24(6):2799–2816, 2008.

## A Signal authentication flow

Figure 4 shows a flow diagram of different screens in Signal when the encryption key changes for a contact, along with the transitions between screens based on user input. The top path is the "message not delivered" flow, which appears to send a message but shows a status indicating that the send failed. The bottom path is the "message delivered" flow, which only shows a notififcation but otherwise proceeds normally. The middle path is the "message blocked flow", which prevents the user from sending a message initially.

## B Redesigned Signal authentication flow

Figure 5 shows the redesigned authentication flow. There is only a single path, using a blocked message dialog along with a shield message in the conversation log.

Figure 6 shows the new notifications that correspond to this flow. If the user attempts to send a message after the encryption keys have changed, the message is blocked and a privacy check dialog is shown (upper left). From here, if the user taps "Get Started", they proceed to the privacy check screen (top middle). They can use either the phone call (top right) or QR code scanner (bottom right). They can choose "Not Now" from either the privacy check dialog or the privacy check screen, and they will proceed to the reminder dialog (bottom left). The result of the privacy check (failure or success) is shown in Figure 7.

Figure 8 shows the notifications in the conversation log. First, when encryption keys change, a notification is displayed that recommends a privacy check (Figure 8a). Later, if the user completes the privacy check, a different notification is shown if the identifiers match (Figure 8b) or don't match (Figure 8c). These notifications scroll as new messages are added to the conversation.

## C Privacy check indicators

Figure 9 shows the new privacy check indicators. Tapping on the indicator brings up the correpsonding privacy check screen, depending on the current state of the conversation, as shown in Figure 10. These same screens are accessed if a user taps of any the conversation log notifications.

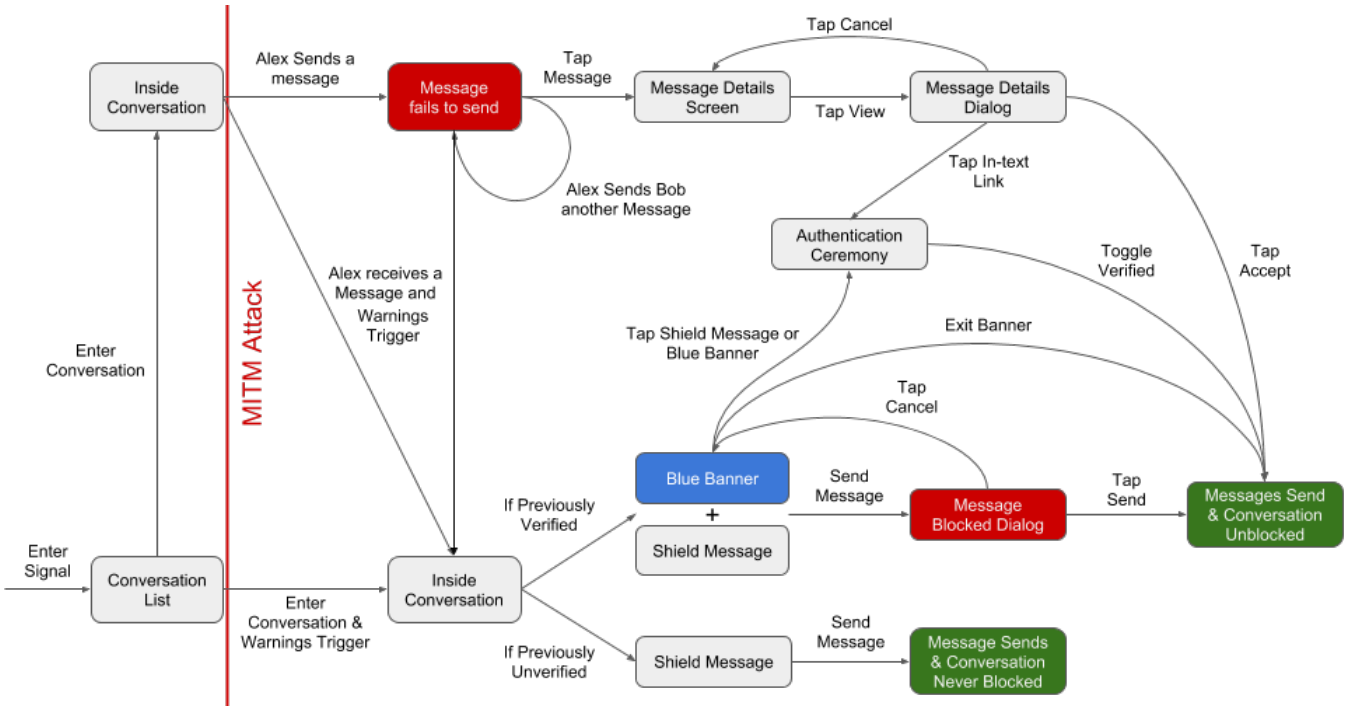



Figure 4: Flow diagram depicting the how Signal reacts to a safety number change.

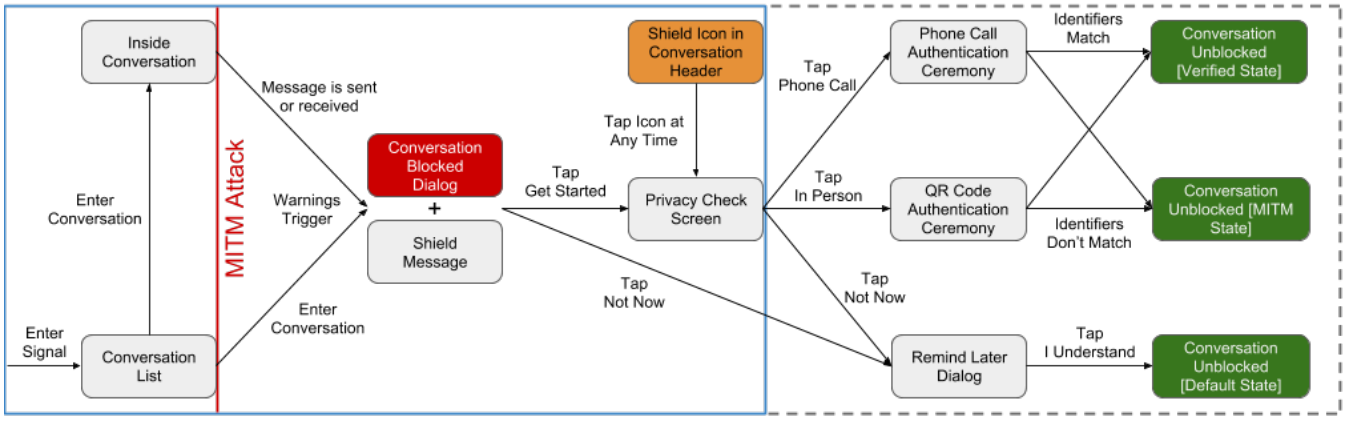
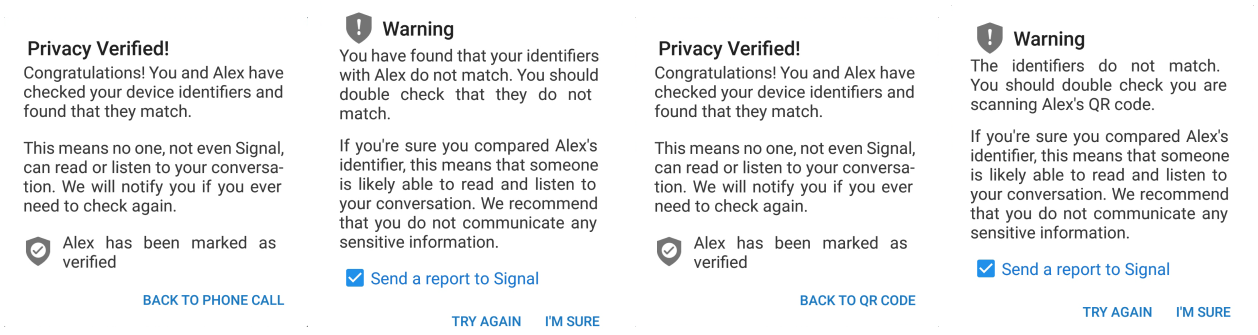



Figure 5: Flow diagram depicting how our redesigned SIgnal reacts to a safety number change. The blue box encloses the elements and choices with analogous equivalents in the original Signal client. The area contained by the dashed lines shows choices, elements, and state changes that we added in our version that are expansions on the authentication ceremony and beyond.


Figure 6: Privacy check notification flow



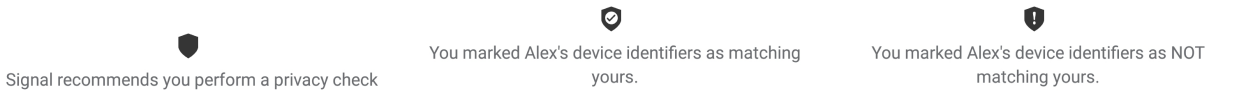

(a) Phone call: matching identifiers

(b) Phone call: non-matching identifiers

(c) QR-code: matching identifiers

(d) QR-code: non-matching identifiers

Figure 7: Phone call and QR code privacy check results




(a) Notification after a key change

(b) Notification after matching identifiers in privacy check

(c) Notification after non-matching identifiers in privacy check

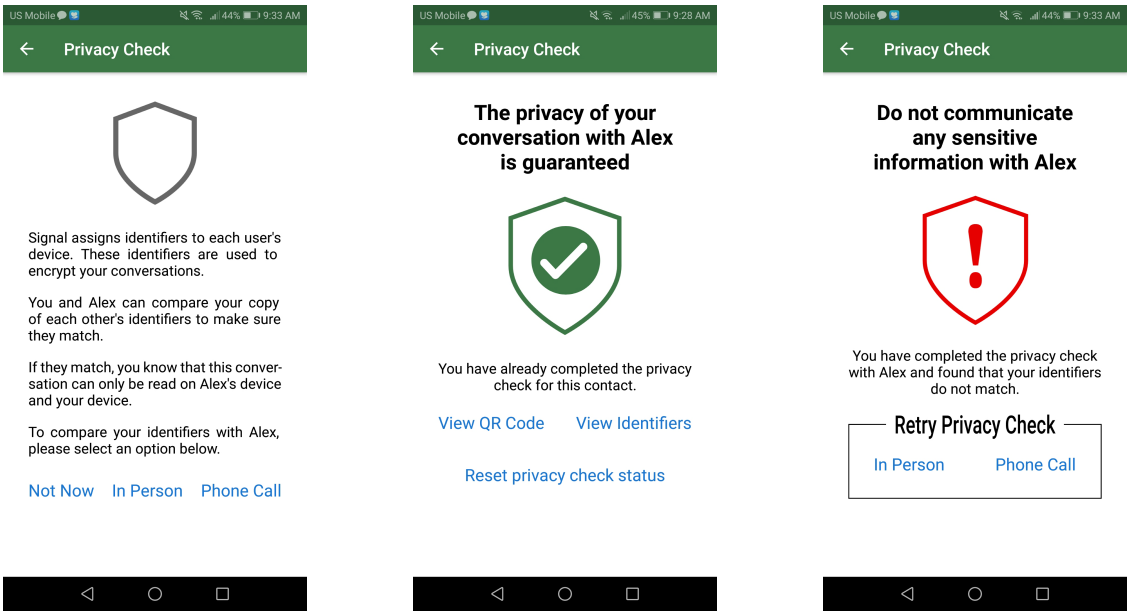
Figure 8: Conversation log notifications


(a) Default state

(b) Matching identifiers

(c) Non-matching identifiers

Figure 9: Privacy check indicator




(a) Default state

(b) Matching identifiers

(c) Non-matching identifiers

Figure 10: Privacy check screen

# "I was told to buy a software or lose my computer. I ignored it": A study of ransomware

Camelia Simoiu
*Stanford University*

Christopher Gates
*Symantec*

Joseph Bonneau
*New York University*

Sharad Goel
*Stanford University*

## Abstract

Ransomware has received considerable news coverage in recent years, in part due to several attacks against high-profile corporate targets. Little is known, however, about the prevalence and characteristics of ransomware attacks on the general population, what proportion of users pay, or how users perceive risks and respond to attacks. Using a detailed survey of a representative sample of 1,180 American adults, we estimate that 2%–3% of respondents were affected over a 1-year period between 2016 and 2017. The average payment amount demanded was $530 and only a small fraction of affected users (about 4% of those affected) reported paying. Perhaps surprisingly, cryptocurrencies were typically only one of several payment options, suggesting that they may not be a primary driver of ransomware attacks. We conclude our analysis by developing a simple proof-of-concept method for risk-assessment based on self-reported security habits.

## 1 Introduction

Ransomware is a particularly pernicious form of malware that restricts an individual's access to their computer (e.g., by encrypting their data) and demands payment to restore functionality. While the first documented ransomware attack dates back to 1989, ransomware remained relatively uncommon until the mid 2000s [26]. Since then, the attack has been automated and professionalized. It is believed to be highly lucrative, with previous damages estimated at hundreds of millions of dollars per year. For example, the damages caused by a single ransomware variant, CryptoWall3, were estimated to be over $320 million in 2015 alone [1].

Consumers are thought to be the most common victims of ransomware [5, 7]. While most attacks are thought to be untargeted, consumers are often less likely to have robust security in place, increasing the likelihood of falling victim to an attack [7]. Despite the harm ransomware can inflict, relatively little is known about the prevalence and characteristics of such attacks in the general population. Reliable estimates of the prevalence of ransomware are necessary both for understanding the nature of today's threat landscape, as well as for longer-term comparison and analysis.

Various government, industry organizations, and researchers have attempted to document the phenomenon, but results have been often inconsistent. This is in large part due to the non-representative data they are based on. Industry reports are typically published by security firms and are based on users of their software products. Such samples are thus inevitably biased towards a set of consumers who have sufficient security awareness and the financial resources to purchase such products. Their experiences may thus not reflect those of the general population. In contrast, government agencies typically report rates based on voluntary victim reports. These estimates are thought to grossly underestimate the true rate [33]. For example, the U.S. Department of Justice estimates that only 15 percent of the nation's fraud victims report their crimes to law enforcement [2], however it is unclear what the true rate of reporting is in the general population.

Apart from the difficulty in characterizing the extent of the problem, little is known about the factors and behavioral patterns that place individuals at risk of such attacks. Devising accurate risk assessment methods to identify the vulnerable population is particularly relevant for ransomware attacks, as infection may impose an especially high cost to consumers. There is often little recourse for victims who need to recover their data other than to pay the ransom. Once identified, information about the vulnerable population can be used to establish proactive strategies to mitigate the effects of ransomware attacks for those individuals that are most at-risk.


*USENIX Symposium on Usable Privacy and Security (SOUPS)* 2019.
August 11–13, 2019, Santa Clara, CA, USA.

For example, the vulnerable population may be influenced through several means, including personalized educational resources and training, or discounted offers for services to mitigate the effects of infection (e.g., cloud-based data backup services). Consumers, if made aware they are at risk of infection, may be better motivated to adopt preemptive measures to mitigate the effects of a potential attack.

We make two key contributions. First, we report the results of a representative online survey of 1,180 U.S. adults that queried respondents' experiences with ransomware attacks. Our results allow us to estimate the prevalence of ransomware in the general U.S. population and responses to such attacks. Second, we develop a simple, proof-of-concept risk assessment for ransomware victimization based on self-reported security habits, which offers an approach for computer users to self-assess their risk of an attack.

## 2 Related work

### 2.1 Estimates of ransomware victimization

The FBI's Internet Crime Complaint Center received 2,673 reports about ransomware in 2016, corresponding to an estimated $2.4 million in losses [4]. The numbers were slightly lower in 2017, with 1,783 ransomware reports received and an estimated $2.3 million in losses [8]. Government estimates, however, are known to underestimate the true rate, as they rely on voluntary self-reports.

Another source of data comes from industry reports that publish experiences of users of their antivirus products. These reports typically use blocked detections as a proxy for actual infections. For example, Symantec reports 405,000 consumer ransomware infections blocked between June 2016 and June 2017 [7]. While analyses by security vendors have the advantage of not relying on self-reports, they suffer from other biases. Industry reports only have visibility into the experiences of the subsection of the population who self-selected to purchase their security product. This sample is likely not representative of the general online population, as it is comprised of individuals who may have a heightened security awareness of online threats, value the product, and have the financial resources to purchase protection. Moreover, blocked detections are imperfect metrics of infection. Traditional, signature-based methods can only detect and block known threats likely missing newer attacks, while more modern machine learning methods suffer from false positives.

More recently, researchers have leveraged public Bitcoin transactions to estimate ransomware infections. For example, Huang et al. [23] provide a lower-bound estimate of 19,750 potential victims globally who made ransom payments using Bitcoins. They do so by (1) scraping reports of ransomware infections in public forums and lists of seed ransom addresses from proprietary sources that maintain a record of ransomware victims and the associated ransom addresses, and (2) extracting ransom addresses by executing several ransomware binaries in a controlled environment. Although the measurement framework presented allows for large-scale measurement of victim rates, it is only able to provide insights into one payment method, namely Bitcoin payments. As we will show, our findings suggest that only focusing on this payment method may provide an incomplete picture of total infections.

### 2.2 Susceptibility to ransomware

The classical paradigm to defend against malware attacks has traditionally been victim-agnostic and reactive, with defenses focusing on identifying the attacks or attackers (e.g., phishing emails, malicious websites, and files) [22]. For example, several studies propose technical, automated solutions to prevent ransomware attacks [16, 25, 26, 35, 39].

More relevant to our work are user studies that identify the vulnerable population and the behaviors that predispose users to malware infections. These cover a wide range of contexts and sub-segments of the population and are typically administered to small, non-representative sample sizes. As a result, it has been difficult to draw conclusions with respect to the general importance of demographic, situational, and behavioral factors on risk of victimization. Ngo et al. [31] apply the general theory of crime and routine activities [17] to assess the effects of individual and situational factors on seven types of cybercrime victimization—among them, a computer virus. They administer an online survey of self-reported cybercrime victimization to 295 students in the U.S., and find that non-white students and younger students had significantly higher odds of obtaining a computer virus. Perhaps counter-intuitively, they also find that individuals who frequently opened any unfamiliar attachments or clicked on web-links in the emails that they received, opened any file or attachment on their instant messengers, and frequently clicked on a pop-up message that interested them, had lesser odds (by about 35%) of obtaining a computer virus.

Bossler et al. [15] conducted a survey of 788 college students to study the risk factors of data loss caused by malware infection. The factors studied include "deviant" behavior (e.g., pirated media downloads, visiting adult websites), routine behaviors (e.g., social media use, programming, shopping), guardianship measures (e.g., having AV software, sharing passwords), and computer skills. The authors find that being employed and being female increased the odds of malware victimization. Engaging in deviant behavior was generally not a strong predictor of malware infection—only pirating media increased the risk of malware infection. Guardianship played small roles in explaining infections, and strong computer skills and careful password management did not reduce estimated threat of malware victimization. Milne et al. [30] conduct a national online survey of 449 US online shoppers. They find that gender, age and number of hours spent online,

excluding email, have a significant impact on users' likelihood to adopt risky online behaviors, concluding that male, younger users, and users who spend many hours online were more at risk.

More recently, researchers have turned to large-scale, data-driven approaches to predict user risk of various cyber threats. Maier et al. [29] examine whether the risk of generating malicious traffic is correlated with security hygiene using DSL data logs of anonymized network traces. They find that having good security hygiene (e.g., applying operating system software updates) has little correlation with being at risk, while accessing blacklisted URLs more than doubles risk.

Levesque et al. [27, 28] observe malware exposure of 50 subjects over a four month period using instrumented computers from the clinical trial of an antivirus product. They find that malware victimization is correlated with a high self-reported level of computer expertise, increased file downloads and application installations, and high browsing volume. The authors find mixed results with respect to the age of the user and the content categories of websites.

Using Symantec telemetry for a subset of 1.6 million users over an 8-month period, Ovelgonne et al. [32] study the relationship between the number of attempted malware attacks detected and user profiles. The authors classify users into 4 categories (gamers, professionals, software developers, others), and find that software developers are more at risk of engaging in risky cyber-behaviors and that there is a sub-population of gamers with especially risky behavioral patterns.

Yen et al. [38] and Bilge et al. [14] study individual user-level malware encounters in an enterprise setting. Yen et al. draw on web proxy logs, user demographics, and VPN logs from a large, multi-national enterprise. The authors investigate features related to categories of web sites visited, aggregate volumes of web traffic, and connections to blocked or low-reputation sites. Using a logistic regression model for inferring the risk of hosts encountering malware, they find that among the three feature categories, user demographics is the strongest indicator of risk, followed by VPN behavior. Counterintuitively, web activity contributed marginally to the overall model and the authors reasoned that this is due to the fact that only 3% of the hosts encountered malware from the web.

# 3 Survey Methodology

## 3.1 Sample selection

We administered a survey on ransomware experiences to a sample of 1,180 U.S. adults. Participants were recruited between June 20, 2017 and September 6, 2017 by YouGov, an online global market research firm, and reimbursed for their participation[1]. YouGov employs a panel of 2 million opt-in participants in the U.S and actively recruits hard-to-reach respondents, such as younger people and those from ethnic minorities, via a network of partners with access to a wide range of online sources that cater to these groups. [2]

In order to derive nationally representative estimates of the U.S. population, YouGov draws stratified samples that approximate the characteristics of random samples of the U.S. population. The sampling frame was designed to match the population in the full 2010 American Community Survey (ACS) conducted by the US Census, and was augmented with voter and consumer databases using the November 2010 Current Population Survey. Summary statistics detailing the demographics of respondents are provided in Table 1, and a more extensive exposition of demographics and socioeconomic characteristics is given in A2 in the Appendix.

## 3.2 Adjustment weights

When constructing a representative sample, non-response and self-selection bias are two common problems that occur, resulting in some population groups being over- or under-represented in the final sample [20, 33]. We use sample weights to address these issues, a standard technique to correct for sample bias. At a high-level, each sample member is assigned a weight such that respondents in under-represented groups receive a weight larger than 1, and those in over-represented groups receive a weight smaller than 1.

YouGov provided adjustment weights for our full sample of 1,180 respondents, which we use throughout our analysis to weight responses.[3] The weights were created by matching to the sampling frame using propensity scores. The matched cases and the frame were combined and a logistic regression was estimated for inclusion in the frame. The propensity score function included age, gender, race/ethnicity, and years of education; propensity scores were then grouped into deciles of the estimated propensity score in the frame and post-stratified according to these deciles.

## 3.3 Defining a ransomware attack

We define ransomware as the class of malware that attempts to defraud users by restricting access to the user's computer or data, typically by locking the computer or encrypting data. There are thousands of different ransomware strains in existence today, varying in design and sophistication [13]. Some ransomware strains can be easily circumvented, while others employ a variety of advanced tactics. For example, they may utilize payload persistence, ensuring the ransomware persists after a restart; use strong encryption methods that are nearly

[1] Participants accumulate points on YouGov for each survey they complete, which an later be redeemed for cash rewards or gift cards at a number of retailers (e.g., Amazon, Best Buy, Target etc.).

[2] The sources include search engine optimization (SEO), affiliate networks, niche websites, and growth hacking techniques such as panelist refer-a-friend campaigns and social networks [6].

[3] Weighted rates were typically quite close to the raw proportions. Omitting the weights did not change the results qualitatively.

| | Raw prop. | Weighted prop. |
|---|---|---|
| Female | 55% | 54% |
| Male | 45% | 46% |
| White | 81% | 75% |
| Black or African American | 8% | 11% |
| Hispanic or Latino | 5% | 8% |
| Asian | 2% | 2% |
| Native American | 1% | 1% |
| Other | 4% | 3% |
| Age (19 – 30) | 11% | 16% |
| Age (31– 45) | 19% | 25% |
| Age (45 – 60) | 28% | 27% |
| Age (over 60) | 42% | 32% |
| Some high school | 1% | 2% |
| High school | 20% | 31% |
| Some college | 22% | 22% |
| College | 39% | 33% |
| Post-graduate | 17% | 12% |

Table 1: *Demographic information and highest level of education achieved (n=1,180 respondents). The raw proportion represents the fraction of respondents and the weighted proportion represents the post-stratified proportion.*

impossible to reverse; or disable system restore functionality (e.g., delete Windows shadow copies) in order to prevent encrypted data from being restored to an older, unencrypted version [19].

Yet another class of ransomware, sometimes referred to as "fake ransomware", informs infected users that their data has been encrypted or their computer locked, however does not actually do these things. These types of attacks are less sophisticated from a technical perspective, are usually relatively easy to circumvent, and rely on scare tactics to coerce the user into paying the ransom amount. We ask respondents to report all types of ransomware, and distinguish between different types, post-response.

## 3.4 Establishing victimization status

Respondents were asked to report any ransomware attack that they had experienced in the past. In order to ensure the accuracy of self-reported ransomware attacks, respondents progressed through a series of ten question and information pages describing typical ransomware attacks and their characteristics. Respondents were initially shown the following definition of ransomware: *"Ransomware is a type of malware that will either lock your computer screen or encrypt your files. If you've been infected with ransomware, you will see screens like the examples below, informing you that you must pay a ransom to re-gain access to your computer and/or files, providing instructions on how to do so."*

Three screenshots of ransomware variants were shown as examples: a strain impersonating the FBI and two encryption ransomware variants, with and without a timer (Figure 1). In


Figure 1: Respondents were shown these three sample screenshots of ransomware: a strain impersonating the FBI and two encryption ransomware variants, with and without a timer.

order to distinguish ransomware attacks from malware with similar characteristics, respondents were shown a page explaining how ransomware is different from technical support scams. [4]

A series of additional questions were then used to confirm respondents' self-reported victimization status. Three multiple choice questions asked whether they experienced various characteristics commonly found in ransomware attacks, namely: (1) whether they had seen similar images notifying them that their computer was locked or files/data encrypted; (2) whether their files were encrypted and they saw files with names such as "DECRYPT INSTRUCTIONS.HTML" or with unusual extensions such as ".locky"; and (3) if they saw a timer counting down and messages indicating that if payment is not completed before the time expires, the ransom amount will increase or the encryption key will be deleted. The ransomware definition above was then repeated and respondents were asked whether they had experienced a ransomware attack, and could answer *Yes*, *No*, or *I am not sure*. If respondents indicated that they were not sure or if their initial response was inconsistent with their answers to the three questions on ransomware characteristics (e.g., they checked off at least one of the three characteristics, but concluded that they did not have ransomware), they progressed through a series of further clarification questions and information pages, which ultimately culminated with the same question (as above) asking them to confirm whether or not they had been infected with ransomware.

Respondents indicating they had experienced a ransomware attack either in the first or second prompt progressed to a series of questions soliciting information about the attack. They

[4] Technical support scams are misleading application that alert the user to a fictitious security issue or vulnerability on their computer, and then prompt them to call a tech support number or to download or purchase anti-virus software in order to resolve the issue.

were asked to describe the ransomware attack in their own words, with prompts to include the contents of the message or instructions, the appearance of the screen, and if any functionality of their computer was disabled. They were also asked a series of questions detailing: the month and year of the attack, the name of the ransomware variant, how much ransom (money) was demanded, the method of payment, whether they paid the ransom and why (or why not), whether access was restored after payment (if applicable), which strategies, if any, they attempted to remove the ransomware and restore access to their computer, whether they sought help in removing the ransomware, whether they were able to remove the ransomware without losing data, how the ransomware was eventually removed, and whether they notified the authorities.

These questions served a dual purpose: apart from allowing us to distinguish between strains with differing attributes (e.g. encryption, screen lock, impersonation of law enforcement), they provided an additional means of validating that the reported incident was indeed a ransomware attack. If respondents had experienced more than one ransomware attack, they were instructed to respond to all questions based on the last attack.

### 3.5 Ethical considerations

All aspects of our study were approved in advance by the university's Institutional Review Board (IRB protocol number 40466). Participants had to reside in the United States and be over 18 years of age to participate. The average completion time was 9.1 minutes (sd=6.8min). Respondents had the option to withdraw at any point during the survey without providing any reason. We informed them that in such a case, none of their data would be used in the analysis. No participant withdrew. Prior to running the study, the survey tool was piloted on Amazon's Mechanical Turk. Five pilot tests with 100 participants each were run. The average completion time ranged from 5min - 8min and each participant was reimbursed the equivalent of $10/hr for completing the survey. Following each pilot, the tool was updated based on preliminary results and feedback from respondents.

### 3.6 Limitations

By design, our survey instrument was intended to mirror the general U.S. population along demographic and socioeconomic dimensions. Nevertheless, it may be still be biased along dimensions other than those that we have explicitly accounted for. For example, YouGov respondents may have technological expertise and privacy and security preferences that differ from those of the broader population.

Additionally, as is the case with nearly all surveys, our results are subject to limitations as they are based on self-reported infection rates rather than upon actual detections of malware. Despite our efforts to ensure that respondents understood what a ransomware attack is, we cannot be certain that ransomware attacks or their attributes were correctly identified. For example, if respondents did not remove the ransomware themselves or did not try relevant troubleshooting strategies (e.g., changing the extensions of files back to the original in order to test whether the encryption was real or not), we cannot know with certainty whether strong encryption was used. Some participants may have confused locked or encrypted files with other problems such as corruption, deletion, or other access issues. Additionally self-reported responses could reflect inaccurate recall and social desirability bias. For example, participants may have been embarrassed to report paying a ransom to restore their data despite their answers being anonymous.

We made several attempts to mitigate these issues. First, the survey tool was piloted on Amazon's Mechanical Turk prior to running the study. The pilots included options to select "I am not sure" or "other" on each question, allowing us to identify which questions were creating confusion as well as any missing options in our multiple choice questions. A follow-up free-text question was displayed to participants that selected "I don't know" to understand the source of confusion, allowing us to further refine our survey tool on each iteration. Second, all self-reports of ransomware victimization for the study on YouGov respondents were independently reviewed by two independent researchers and re-classified when necessary (details provided in Section 4.1).

One limitation to piloting on Mechanical Turk is that the population there differs from that used for the final study. Mechanical Turkers, for example, have been found to be a relatively tech-savvy group [33] and YouGov respondents may not have interpreted questions in the same way due to lower technical expertise. To mitigate these issues, future work could pilot the survey tool on the same population as the target population. The use of focus group with members of the public would also allow researchers to ask follow-up questions on the survey and discuss any ambiguities with respondents. While our results may be impacted by these limitations, we believe that our study is still a step forward in understanding ransomware experiences for the general online population.

## 4 Results

### 4.1 Re-classifying victimization status

All responses from self-reported ransomware victims were independently reviewed by two independent researchers, and all conflicting classifications were reviewed and resolved. Each response was classified under two regimes: a *conservative* regime and an *inclusive* regime. [5] Both regimes exclude cases where the respondent described a different type of malware

[5] Cohen's kappa measuring inter-rater agreement prior to reaching consensus was 0.53 and 0.66 for the conservative and inclusive regimes, respectively.

attack (e.g., scareware, pop-up announcing that they were the winner of a contest, or tech support scam), or admitted that they did not remember the details of the attack. The difference between the two regimes is relevant for ambiguous cases. The conservative regime includes only cases where the description of the attack provides sufficient information to confirm beyond reasonable doubt, that it was ransomware. The inclusive regime includes, in addition to the above, cases where the description was ambiguous or no description was provided. We include both regimes as both require an assumption on the part of the coders, namely either that the respondent understood what a ransomware attack was and simply did not choose to provide lengthy description (inclusive regime), or did not understand what a ransomware attack was (conservative regime). Fortunately, estimated prevalence was similar under both the inclusive and conservative classification schemes. For this reason, with the exception of the ransomware rate, all results are reported for victims classified under the *inclusive* regime for ease of exposition. Sample responses and their classification under the two regimes are listed in Table 2.

Prevalence is estimated by the following:

$$r = \frac{\sum_i I_i(\text{victim}) w_i}{\sum_i w_i} \tag{1}$$

where $I_i(\text{victim})$ is an indicator function representing whether the respondent reported experiencing ransomware or not, and $w_i$ is the weight.

## 4.2 Rate of ransomware victimization

Originally, 153 respondents (14%) reported that they had experienced a ransomware attack at some time in the past (Table 3). Following re-classification, we estimate that the overall proportion of the U.S. population reporting a ransomware infection at any time in the past ranges between 6% (se=1%, n=63) under the conservative regime, and 9% (se=1%, n=96) under the inclusive regime[6] We similarly estimate that between 2% (se=0.4%, n=19) and 3% (se=0.5%, n=33) of the U.S. population were affected over the one-year period between June 2016 to June 2017 under the conservative, and inclusive regimes, respectively.

Given that there are approximately 200 million U.S. adults who have a computer with internet access [10, 12, 34], our results suggest that several million Americans were victims of ransomware in the 1-year period we studied. We note, however, that multiple people may share the same computer. As such, the number of *victims* of ransomware attacks may be substantially larger than the number of *households* that experienced an attack or the number of infected *computers*.

It is difficult to directly compare our estimates to those from previous studies, but our results appear to be broadly consistent with past evidence. For example, Symantec reported 405,000 consumer ransomware infections were blocked between June 2016 and June 2017 [7]. Given that approximately 25 million U.S. consumers use Symantec AV [11], that suggests an infection rate of 1.6%.[7] Our estimate also appears to be approximately consistent with evidence provided by Huang et al. [23], though we must make several assumptions to compare their reported numbers to our own. Specifically, Huang et al. identify approximately 20,000 bitcoin payments for ransomware globally over a 22-month period, or roughly 10,000 over 12-months. This number, though, likely substantially undercounts the true number of payments because only a fraction of all such bitcoin transactions could be identified, as described in Section 2. In our survey, only 1% of ransomware victims made a bitcoin payment—the vast majority did not pay at all. If that number is representative, the 10,000 bitcoin payments globally translates to approximately 10,000 / 1% = 1 million global ransomware incidents. We caution that one cannot directly compare this estimate to our own: we consider only U.S. victims, not global victims; and the number of bitcoin payments identified by Huang et al. is an underestimate of the actual number of payments. Nonetheless, despite being based on quite different methodologies, the two approaches yield estimates of the same order of magnitude.

## 4.3 Ransomware attributes

We now turn our focus to examining the characteristics of ransomware attacks experienced by respondents. Overall, ransomware strains that lock the computer appear to be more common than those that employ encryption — 74% of victims reported experiencing computer locks, while only 35% reported that their files were encrypted. Our finding is in line with a recent report by Kaspersky Lab, which finds that 40% of users are attacked with encryption ransomware as a proportion of users attacked with any kind of ransomware in the U.S. between 2015 and 2016 [3]. Figure 2 shows the distribution of attributes experienced by ransomware victims. Two observations stand out. First, a large proportion of victims reported experiencing strains that impersonate law enforcement agencies, typically the FBI (46%). These strains typically display a message claiming that the user's computer was locked because they engaged in illegal activities (e.g., browsed illegal pornographic websites), and a fine must be paid to regain access. Second, encryption does not appear to be commonly used in conjunction with law enforcement impersonation. Only 22% of victims reporting law enforcement strains also reported that their files were encrypted, whereas 43% of victims that did not experience law enforcement strains experienced encryption.

[6] Standard errors (se) are given in parenthesis.

[7] To estimate the number of U.S. Symantec users, we scaled the reported number of global Symantec users (50 million) by the reported share of revenue generated by U.S. users (52%).

| Respondent's description | Screen lock | Encryption | Law enforcement | Timer | Inclusive | Conservative |
|---|---|---|---|---|---|---|
| "Illegal files detected. FBI has locked your computer. purchase a prepaid VISA and pay fine online . (included a fake web cam window) | • | | • | | Ransomware | Ransomware |
| "FBI - YOU HAVE BEEN WATCHING PORN OR GAMBLING OR BOTH, YOU MUST PAY $200 TO MONEYGRAM" | • | • | • | | Ransomware | Ransomware |
| "The screen looked like one you previously displayed. It encrypted just about all my files except *.exe files and a few others. I lost everything on my PC and external hard drive. I ended up reformatting and starting from scratch. I could run programs, but could not access any of my saved working files. A screen would display telling me to call a number, pay the ransom and they would decrypt my files. They wanted $500." | | • | | | Ransomware | Ransomware |
| "i don't remember what the message said it just prevented me from getting to any of the stuff on my computer and then i started it in safe mode and got rid of it" | • | | | | Ransomware | False Positive |
| " It wasn't a specific ransom note but it was inferred that unless I bought the software my files wouldn't be at my disposal. They were." | | • | | • | Ransomware | False Positive |
| "It popped up and stated that I had to pay to gain access back to my computer and I was unable to do anything." | • | | | | Ransomware | False Positive |
| "the screen was flashing call this number immediately to get your computer repaired. I was gullible and scared so I called. the guy got a hold of my computer and then told me I had to pay $300 for him to fix it. I told him I didn't have that kind of money and he hung up on me. I then went and changed all my passwords and prayed he didn't get any important info from me." | | • | | | False positive | False positive |
| "I don't recall exactly, However when I called to find out what was going on, I was told that I would have to pay to get what ever was holding up my computer off, I said, You put it on, just take it off. My computer was older, I just went and bought another computer, I decided not to be an ATM for criminals." | | | | | False positive | False positive |
| "Was told to send $1000.00 dollars to clean up computer." | | | | | False positive | False positive |

Table 2: *Sample descriptions and reported characteristics of the attack, and their corresponding classification under the conservative and inclusive classification regimes. Responses classified as false positives under the conservative regime, but not the inclusive regime, are typically classified as such due to unclear or ambiguous descriptions. Responses classified as false positives under both regimes are typically classified as such because the descriptions provided typically describe other scams (technical support scams, scareware, etc.) and they include few ransomware characteristics (if any).*

### 4.4 Ransom payment

A histogram of reported ransomware amounts demanded is shown in Figure 3. The median and average reported ransom is $250 and $530 (standard error $125), respectively, while the maximum amount reported reached $8,000. This finding is approximately in line with Industry reports. For example, Symantec reports the average ransom 2016 to be $1,077 in 2016 and $522 in 2017 [7, 9].

The most common payment methods reported were wire transfers and payment voucher systems (e.g., Paysafecard, MoneyPak, CashU, MoneXy, prepaid Visa)[8], which together accounted for 56% of all reports. In contrast, only 12% of

[8]Respondents were presented with a multiple choice question and asked what payment method they were asked to pay the ransom in. As respondents could not select multiple payment methods, we are able to estimate a lower bound on the distribution of payment methods. Only respondents that were able to recall the ransom amount are included (n=66).

| | All victims | Last 12 months |
|---|---|---|
| Self-reported | 14% | 5% |
| Re-classified (inclusive) | 9% | 3% |
| Re-classified (conservative) | 6% | 2% |

Table 3: *Proportions of ransomware victimization for the U.S. population under the conservative and inclusive classification schemes. The "all victims" column includes respondents who reported experiencing a ransomware attack at any time in the past; the "last 12 months" column includes respondents who reported attacks within one year of the survey date.*

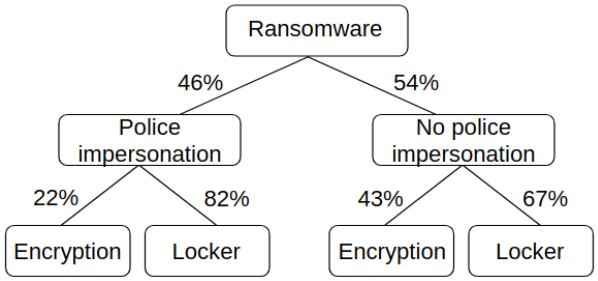



Figure 2: *Distribution of ransomware attributes (impersonation of law enforcement, locker, encryption). Categories are not mutually exclusive. Strains employing police impersonation tactics are widespread (46%) and tend to favor locking mechanism as opposed to encryption.*

respondents reported being asked to pay only via cryptocurrency. Table 4 shows the distribution of payment method for respondents reporting a ransomware attack[9].

Our results are primarily driven by the predominance of locker ransomware in our sample (recall that 74% of victims reported experiencing locker ransomware). This is consistent with characteristics of ransomware samples observed in the wild. Since ransomware strains that rely on locking techniques (as opposed to encryption) effectively restrict functionality to the computer, they must rely on pre-paid cash vouchers or wire transfers for payment. Encryption ransomware strains typically do not restrict functionality and tend to favour cryptocurrency payment schemes [13].

While recent work has focused on tracking Bitcoin payments as a means to estimate ransomware infections and quantify financial losses [23], this finding suggests that focusing solely on cryptocurrencies may underestimate losses as it focuses on only one of many types of ransomware families. Secondly, it casts doubt on the hypothesis that increased adoption of cryptocurrencies is a main driving force of the recent ransomware trend.

[9]Results are qualitatively similar for victims reporting experiencing a ransomware infection within the last 12 months: 62% reported wire transfers or payment voucher systems whereas only 2% reported cryptocurrencies.

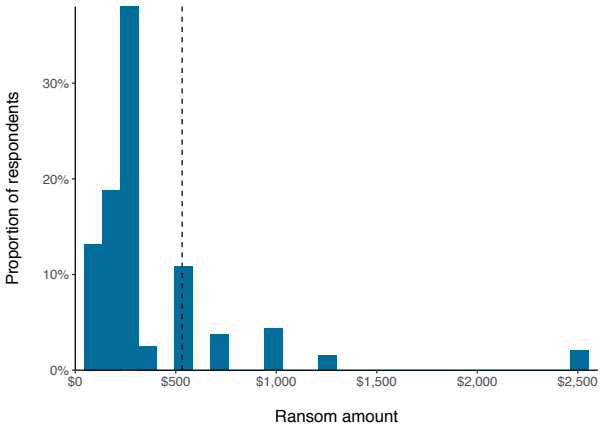



Figure 3: *Histogram of reported ransom amounts for respondents recalling an amount (n=66). The median and mean reported ransom is $250, and $530 (se. $125), respectively. The dashed line represents the mean. The maximum amount reported, $8,000 is omitted from the plot.*

| Method of payment | Proportion |
|---|---|
| Pre-paid cash voucher | 42% |
| Wire transfer | 14% |
| Cryptocurrency | 12% |
| Premium-rate text message | 7% |
| Not displayed | 15% |
| Do not remember | 10% |

Table 4: *Distribution of payment methods. Wire transfers and pre-paid cash vouchers predominate, whereas cryptocurrencies account for 12% of reported payment methods. The"not displayed" category includes cases where the payment method was not directly displayed (e.g., respondents would have had to follow a link to find out, and did not do so).*

## 4.5 Means of dealing with the attack

Respondents reported a wide range of methods for dealing with the ransomware infection, depending on the severity of the strain. As detailed in Table 5, the majority of victims either found a tool online to remove the ransomware and/or decrypt their files, restored their computers from backups, or re-started their computers (e.g., in safe mode). 13% of victims obtained help in removing the ransomware, either paying for repairs at a computer shop or asking friends or family for help.

Few victims paid the ransom (n=6, 4%) or reported the attack to authorities (n=7, 9%). Access was restored for all victims that did pay. The self-reported reasons for paying the ransom focused on feelings of distress, aversion to losing files, as well as lack of computer knowledge. Respondents' original reasons are given below:

| Method | Proportion |
|---|---|
| Restarted computer | 30% |
| Online tool | 18% |
| Restored computer from backup | 22% |
| Removed by someone else | 13% |
| Reformatted computer | 5% |
| Removed using AV software | 5% |
| Paid ransom | 4% |
| Other means | 3% |

Table 5: *Self-reported means of dealing with the attack.*

1. *"I am computer illiterate. A little smarter now."*
2. *"We were very distressed and felt it was a legitimate request."*
3. *"Did not want to lose any files or programs on my system."*
4. *"I'm so scared"*
5. *"I was a full time caregiver for my critically ill husband. He used the computer a great deal to maintain contact with friends and family. I did not want to take the computer somewhere to have the problem corrected at what likely would have been a more expensive cost.*
6. *"The price was not that high."*

To note is that financial losses associated with paying the ransom only capture one dimension of the total costs imposed on victims. These include psychological costs associated with losing valuable data (e.g., family photos) and time costs of dealing with the aftermath of the attack. As one respondent details, *"It was a mess for a while [...] and very troubling, my husband worked on it for a whole day."* In addition, victims may incur additional financial costs to deal with the attack such as paying technicians to remove the ransomware or investing in protection tools such as anti-virus products to prevent future infections.

## 4.6 Behavioral changes post-attack

Victims were asked to indicate whether they changed any of their habits following the attack, if any. We find that 56% of respondents reported changing two or more habits. The top three changes reported were more careful browsing (65%), purchasing an antivirus software (44%), and updating their existing antivirus product (31%) (Table 6).

Few respondents reported changing their operating system (OS), although we find that victimization varies significantly with OS. We find that 10% of Windows users were victims, whereas only 5% of non-Windows users were victims. This difference is statistically significant using a two-proportion Z-test at the 5% significance level. The majority of respondents used Windows as their OS (82%)[10]. Only 26% of respondents began to backup their data or backed up their data more frequently following the attack.

| Habit | Proportion |
|---|---|
| More careful browsing | 65% |
| Purchased AV product | 44% |
| Updated AV product | 31% |
| Started to backup data | 26% |
| Enable automatic updates | 24% |
| Backup data more regularly | 22% |
| Changed OS configurations | 20% |
| Changed OS | 10% |
| Changed default browser | 12% |
| Encrypted hard drive | 0% |

Table 6: *Behavioral changes following the attack for ransomware victims. Multiple answers were permitted. The top three changes reported were more careful browsing, and purchasing or updating an antivirus product. "Enable automatic updates" refers to updates to the OS, browser, antivirus, and other programs. Examples of configuration changes are disabling Windows Script Host, restricting login access, enabling the "show file extension" feature in Windows.)*

Whether or not participants truly changed their habits following the attack, or if this is a form of social desirability bias, is difficult to know for sure. Nevertheless, two observations stand out. First, this result suggests that the majority of victims attribute the cause of the attack, at least in part, to their own behaviors. At the very least, they display the intention to change their behaviors in order to minimize their risk. Secondly, data backup habits are arguably the single most effective way to mitigate the effects of ransomware attacks, yet few respondents adopt this behavior even after experiencing an attack. This suggests that more awareness is needed around the importance of this habit.

## 4.7 Perceptions of risk and responses

Along with precautionary security habits and online behaviors, risk perception — or the awareness of one's susceptibility to adverse security outcomes — is thought to play an important role in making better security decisions [37]. We investigate how experiencing a ransomware infection affects perception of risk via two questions: (1) "How likely do you think you are to experience a ransomware attack in the future?" and (2) "Suppose you were to experience a ransomware attack today and the only way of restoring access to the data on your computer was to pay the ransom (say $300). How likely is it that you'd pay the ransom?". Participants were prompted to enter a number between 0 and 100, where 100 means: "I'm

[10] 12% used a Mac, 4% used Chrome, while the remaining 2% used another OS.

definitely (100% likely) going to [experience a ransomware attack in the future / pay the ransom]." and 0 means: "There is no way (0% chance) I will [experience a ransomware attack in the future / pay the ransom]".

Whereas victims reported a mean of 47 (sd=34) for the likelihood of experiencing a future ransomware attack, non-victims reported a mean of 30 (sd=25). This difference was significant based on an independent-samples t-test, t (104) = -4.97, 95% CI of the difference (10.78, 25.07) Similarly, victims reported a mean=2.9 (sd=11) for the likelihood of paying the ransom, versus a mean of 8.4 (sd=20) for non-victims. The difference was significant: t (158)=4.26, 95% CI of the difference (2.93, 8.00). These results suggest that victims believe they are more at risk of a future attack, and less likely to pay a ransom. This may be due to victims feeling better prepared to deal with a future attack due to a change in habits or improved mitigation strategies, or feeling less uncertainty about the consequences of an attack after having experienced one. Further research, however, is needed to understand the exact reasons for these differences and carefully mitigate any response biases that may exist here.

## 5 Predicting ransomware infection

Given the potentially high cost of a ransomware infection, a natural follow-up question is whether it is possible to identify the set of at-risk users. Once identified, the hope is that we can mitigate the effects of an infection for those individuals that are most likely to experience an attack. In the same vein, employers could offer personalized educational resources and training; antivirus companies could fine-tune and re-prioritize defense mechanisms to offer additional protection layers, set different default settings, or partner with vendors to provide discounted offers for services to mitigate the effects of infection (e.g., online backup services). Finally, consumers—if made aware they are "at risk"—may be better motivated to improve their security posture and adopt better security habits. For example, in several health domains, Strecher et al. [36] found that perceived susceptibility—the belief that one is at risk for the issue at hand—was a necessary factor to achieve behavior change.

### 5.1 Traditional, machine-learned models

To estimate risk of infection, we start by training traditional statistical models on our survey data. We consider the complete set of responses (n=1,180) and define positive examples to be those that have experienced ransomware at any time in the past (9%, n=96). Given each respondent's answers, we construct a model to predict infection status using two standard machine learning models: lasso (a linear model), and gradient boosted trees (GBM, a non-linear model). To do so, we draw on several features extracted from the survey: demographics, socioeconomic status, the software used, level of computer knowledge[11], and general security habits. Table A1 in the Appendix includes a comprehensive record of features extracted from the survey questions, several of which have been inspired by previous work [15, 18, 29–31].

| Features | Lasso | GBM |
|---|---|---|
| Dem + SES | 65 | 63 |
| Dem, SES, Tech, Computer | 61 | 65 |
| Habits | 66 | 67 |
| Habits + Scam | 75 | 74 |
| All features | 76 | 76 |

Table 7: *Average AUC across K=10 folds for lasso and gradient boosting tree (GBM) models using demographics ("Dem"), socioeconomic covariates ("SES"), the technology used ("Tech"), computer knowledge ("Computer"), security habits ("Habits"), and an indicator of previously experienced an online scam ("Scam"). Models based solely on self-reported security habits and previous experience with online scams performed on par with the saturated models using all covariates.*

We believe these features are appropriate to illustrate the general predictive power of such information. But we suggest that future work along these lines make use of scales that have been expressly designed and validated to measure the relevant information. Doing so can lead to a more accurate measurement of underlying behaviors, and may ultimately lead to improved predictive performance.

We evaluate our predictive models with stratified K-fold cross-validation, where K=10, [12] and report performance in terms of average AUC score across the folds, in Table 7. We find that models using only demographic and socioeconomic features achieve a maximum average AUC of 65%. Slightly higher performance is achieved using only features related to security habits (67% average AUC). Previously experiencing an online scam also proves to be highly predictive of ransomware infection, and the model including both security habits and past experience with an online scam achieves performance on par with the saturated model that includes all features (an average AUC of 75%).

### 5.2 A simpler approach to risk assessment

Given the results above, we now present and discuss a proof-of-concept approach to risk assessment to estimate future ransomware infection that is based only on self-reported security habits and past exposure to online scams. The method demonstrates that assessments can, in theory, be made with

[11] We assess the level of computer knowledge using an 8-question test developed by the authors.

[12] The data is randomly partitioned into K=10 equal sized subsamples with the proportion of positive examples equal to that in the full data set. A single subsample is retained as the validation data for testing the model, and the remaining K - 1 subsamples are used as training data.

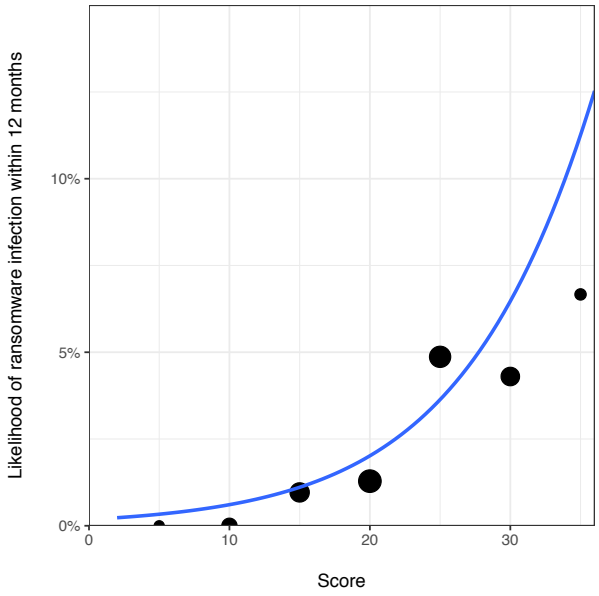



Figure 4: *Calibration plot for our simple approach to risk assessment, showing the calculated score versus the empirical proportion of ransomware infections within 12 months, and the fitted logistic regression line. Scores are grouped into buckets of 5, with each bucket containing between 3 to 13 infections. Higher scores correspond to an increased likelihood of infection.*

relatively little information, enabling consumers to estimate their own risk. We stress from the outset, however, that we merely intend to illustrate the general approach; in particular, the strategy we present would need to undergo more rigorous evaluation before it could be responsibly used for risk assessment in the broader population.

Following Jung et al. [24], we use the "select-regress-and-round" method to create a weighted risk-assessment rubric, which we find performs on par with the traditional machine learning algorithms described above. The rubric is constructed using the output from the tuned lasso model presented in Section 5.1, where we re-scale and round the resulting model coefficients to yield integer weights from 1 to 10 [24].[13]

The final risk assessment rubric is based on six factors: use of two-factor authentication, data backup habits, encryption of hard drive, frequency of using torrent services, password protection for login, and previous experience with online scams. The complete list of questions used to assess risk of ransomware, and their corresponding scores, are included in Table 8. Higher scores correspond to a higher likelihood of infection. We find that this simple approach to risk assessment performs on par with more complex models, achieving average cross-validated AUC of 78% across $K = 10$ folds.

[13]The coefficients are normalized to integers on a scale of 1-10, where the scaled coefficients are equal to $c_{\text{scaled}} = \text{round}(c_{\text{original}} * 10/c_{\max})$ and $c_{\max}$ is the maximum coefficient produced by the original model. Questions with re-scaled coefficients that round to zero are dropped from the rubric. For each remaining question, the re-scaled coefficient is multiplied by each possible answer to obtain the points in Table 8.

| Question | Points |
|---|---|
| How frequently do you download files from online torrent sites such as the Pirate Bay, ExtraTorrent, or TorrentZ2? | |
| • I frequently download files from torrent sites. | 15 |
| • I occasionally download files from torrent sites. | 10 |
| • I rarely download files from torrent sites. | 5 |
| • I never download files from torrent sites. | 0 |
| Do you backup your personal files to an external hard drive or a cloud-based storage service? | |
| • I do not have any of my files backed up. | 8 |
| • I backup my files once a year. | 6 |
| • I backup my files every couple of months. | 4 |
| • I backup my files every couple of weeks. | 2 |
| • I backup my files every day. | 0 |
| Is your hard drive encrypted? | |
| • Yes, my hard drive is encrypted. | 0 |
| • No, my hard drive is not encrypted. | 1 |
| Have you ever downloaded—or been asked to download—an application that you suspect was malicious, like fake anti-virus software? | |
| • Yes, I have. | 10 |
| • No, I haven't. | 0 |
| Do you use two-step authentication for at least one of your online personal accounts (i.e., not for a work-related account)? | |
| • Yes, I use two-step authentication. | 0 |
| • No, I don't use two-step authentication. | 1 |
| Is your computer password-protected for login? | |
| • Yes, my computer has a password. | 0 |
| • No, my computer doesn't have a password. | 8 |

Table 8: *Questions included in our simple risk assessment rubric based on self-reported security habits and previous experience with online scams.*

To aid interpretation, we convert risk scores to probability of infection as follows: we first calculate the risk score for each respondent using the derived weights in Table 8, and then predict ransomware status within 12 months via logistic regression using the calculated risk score as the sole feature. In Figure 4, we show the resulting calibration plot for the risk scores. For example, a risk score of 15 corresponds to 1% likelihood of infection.

It bears emphasis that the risk assessment method we present is only *predictive*, in the sense that the factors we identify are *correlated* with the risk of infection; the features we use are not necessarily *causally* related to future infection. For example, not backing up your data is correlated with infection, although opting to regularly back up your data will

not cause the likelihood of infection to decrease. Further, the relationship between the predictive factors we identify and ransomware infection will likely change over time. For example, as it becomes easier and less expensive to backup data, doing so may be less indicative of technical savviness and, accordingly, may be less predictive of ransomware infection. Finally, we have carried out our analysis on a relatively small dataset of users.

## 6 Conclusions and future work

Our survey results shed new light on the scale of ransomware in the general population and the actions users took in response. Our estimated victimization rate of 2–3% of the population per year suggests millions of ransomware cases per year. An important future research question is whether these figures are growing (and at what rate), which will require longitudinal follow-up studies.

Conventional wisdom has held that cryptocurrencies would fuel growth in ransomware, but our results suggest most cases in 2016–2017 were not reliant solely on cryptocurrency for payment. Another open question for future research is if payment rates will increase or decrease as more individuals affected have either been previously victimized themselves or have heard more about ransomware from affected friends and family. Follow-up work might study what factors affect payment rates in more detail, how users perceive their susceptibility to attack, what affects their risk perceptions, whether they are well-calibrated, and how previous infections affects their perceptions.

Finally, the simple approach to risk assessment that we present suggests that vulnerability can, in theory, be estimated from self-reported security habits and previous exposure to online scams. Our model is relatively straightforward and transparent, enabling consumers to estimate their own risk of infection. While prior research suggests these qualities make risk-assessments more acceptable to users [21], future research is required to gauge user reaction.

## Acknowledgments

We thank Ansh Shukla, Leyla Bilge, Petros Efstathopoulos, Dan Boneh, and Darren Shou for helpful comments and feedback.

## References

[1] Lucrative ransomware attacks: Analysis of the cryptowall version 3 threat. Technical report, Cyber Threat Alliance, 2015 (accessed August 24, 2018). https://www.cyberthreatalliance.org/resources/lucrative-ransomware-attacks-analysis-cryptowall-version-3-threat/.

[2] Financial crime fraud victims. Technical report, The United States Attorney's Office, Western District of Washington; United States Department of Justice, 2015 (accessed October 12, 2018). https://www.justice.gov/usao-wdwa/victim-witness/victim-info/financial-fraud.

[3] KSN report: Ransomware in 2014-2016, kaspersky lab. Technical report, Kaspersky Lab, 2016. https://media.kasperskycontenthub.com/wp-content/uploads/sites/43/2018/03/07190822/KSN_Report_Ransomware_2014-2016_final_ENG.pdf.

[4] 2016 internet crime report. Technical report, Internet Crime Complaint Center, Federal Bureau of Investigation, 2016 (accessed August 7, 2018). https://pdf.ic3.gov/2016_IC3Report.pdf.

[5] KSN report: Ransomware in 2014-2016. Technical report, Kaspersky Lab, 2016 (accessed January 12, 2019). https://media.kasperskycontenthub.com/wp-content/uploads/sites/43/2018/03/07190822/KSN_Report_Ransomware_2014-2016_final_ENG.pdf.

[6] The YouGov online panel. Technical report, YouGov, 2017 (accessed August 19, 2018). https://d25d2506sfb94s.cloudfront.net/r/93/YouGov_-Online_Panel_Book2017.pdf.

[7] 2017 internet security threat report, symantec, vol. 22. Technical report, Symantec, 2017 (accessed August 6, 2018). https://www.symantec.com/content/dam/symantec/-docs/reports/istr-22-2017-en.pdf.

[8] 2017 internet crime report. Technical report, Internet Crime Complaint Center, Federal Bureau of Investigation, 2017 (accessed January 12, 2019). https://pdf.ic3.gov/2017_IC3Report.pdf.

[9] 2018 internet security threat report, symantec, vol. 23. Technical report, Symantec, 2018 (accessed August 6, 2018). https://www.symantec.com/content/dam/symantec/docs/reports/istr-23-2018-en.pdf.

[10] Internet/broadband fact sheet. Technical report, Pew Research Center, 2018 (accessed January 19, 2019). https://www.pewinternet.org/fact-sheet/internet-broadband/.

[11] Corporate fact sheet. Technical report, Symantec Corporation, 2018 (accessed March 3, 2019). https://www.symantec.com/content/dam/symantec/docs/other-resources/symantec-corporate-fact-sheet-060517-en.pdf.

[12] Annual estimates of the resident population by sex, age, race, and hispanic origin for the united states and states: April 1, 2010 to july 1, 2017. Technical report, U.S. Census Bureau, Population Division, Release Date: June 2018 (accessed August 24, 2018). https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk.

[13] Pranshu Bajpai, Aditya K Sood, and Richard Enbody. A key-management-based taxonomy for ransomware. In *2018 APWG Symposium on Electronic Crime Research (eCrime)*, pages 1–12. IEEE, 2018.

[14] Leyla Bilge, Yufei Han, and Matteo Dell'Amico. Riskteller: Predicting the risk of cyber incidents. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 1299–1311. ACM, 2017.

[15] Adam M Bossler and Thomas J Holt. On-line activities, guardianship, and malware infection: An examination of routine activities theory. *International Journal of Cyber Criminology*, 3(1), 2009.

[16] Krzysztof Cabaj and Wojciech Mazurczyk. Using software-defined networking for ransomware mitigation: the case of cryptowall. *IEEE Network*, 30(6):14–20, 2016.

[17] Kyung-shick Choi. Computer crime victimization and integrated theory: An empirical assessment. *International Journal of Cyber Criminology*, 2(1), 2008.

[18] Serge Egelman and Eyal Peer. Scaling the security wall: Developing a security behavior intentions scale (sebis). In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 2873–2882. ACM, 2015.

[19] Alexandre Gazet. Comparative analysis of various ransomware virii. *Journal in computer virology*, 6(1):77–90, 2010.

[20] Andrew Gelman, Sharad Goel, Douglas Rivers, David Rothschild, et al. The mythical swing voter. *Quarterly Journal of Political Science*, 11(1):103–130, 2016.

[21] G. Gigerenzer, R. Hertwig, and T. Pachur. *Heuristics: The Foundations of Adaptive Behavior*. OUP USA, 2011.

[22] Hassan Halawa, Konstantin Beznosov, Yazan Boshmaf, Baris Coskun, Matei Ripeanu, and Elizeu Santos-Neto. Harvesting the low-hanging fruits: defending against automated large-scale cyber-intrusions by focusing on the vulnerable population. In *Proceedings of the 2016 New Security Paradigms Workshop*, pages 11–22. ACM, 2016.

[23] Danny Yuxing Huang, Damon McCoy, Maxwell Matthaios Aliapoulios, Vector Guo Li, Luca Invernizzi, Elie Bursztein, Kylie McRoberts, Jonathan Levin, Kirill Levchenko, and Alex C Snoeren. Tracking ransomware end-to-end. In *Tracking Ransomware End-to-end*, page 0. IEEE.

[24] Jongbin Jung, Connor Concannon, Ravi Shroff, Sharad Goel, and Daniel G Goldstein. Simple rules for complex decisions. 2017.

[25] Amin Kharraz, Sajjad Arshad, Collin Mulliner, William K Robertson, and Engin Kirda. Unveil: A large-scale, automated approach to detecting ransomware. In *USENIX Security Symposium*, pages 757–772, 2016.

[26] Amin Kharraz, William Robertson, Davide Balzarotti, Leyla Bilge, and Engin Kirda. Cutting the gordian knot: A look under the hood of ransomware attacks. In *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*, pages 3–24. Springer, 2015.

[27] Fanny Lalonde Levesque, Jude Nsiempba, José M Fernandez, Sonia Chiasson, and Anil Somayaji. A clinical study of risk factors related to malware infections. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, pages 97–108. ACM, 2013.

[28] Fanny Lalonde Lévesque, José M Fernandez, and Anil Somayaji. Risk prediction of malware victimization based on user behavior. In *Malicious and Unwanted Software: The Americas (MALWARE), 2014 9th International Conference on*, pages 128–134. IEEE, 2014.

[29] Gregor Maier, Anja Feldmann, Vern Paxson, Robin Sommer, and Matthias Vallentin. An assessment of overt malicious activity manifest in residential networks. In *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*, pages 144–163. Springer, 2011.

[30] George R Milne, Lauren I Labrecque, and Cory Cromer. Toward an understanding of the online consumer's risky behavior and protection practices. *Journal of Consumer Affairs*, 43(3):449–473, 2009.

[31] Fawn T Ngo and Raymond Paternoster. Cybercrime victimization: An examination of individual and situational level factors. *International Journal of Cyber Criminology*, 5(1), 2011.

[32] Michael Ovelgönne, Tudor Dumitraş, B Aditya Prakash, VS Subrahmanian, and Benjamin Wang. Understanding the relationship between human behavior and susceptibility to cyber attacks: a data-driven approach. *ACM*

*Transactions on Intelligent Systems and Technology (TIST)*, 8(4):51, 2017.

[33] Elissa M Redmiles, Sean Kross, Alisha Pradhan, and Michelle L Mazurek. How well do my results generalize? comparing security and privacy survey results from mturk and web panels to the us. Technical report, 2017.

[34] Camille Ryan. Computer and Internet Use in the United States: 2016.

[35] Nolen Scaife, Henry Carter, Patrick Traynor, and Kevin RB Butler. Cryptolock (and drop it): stopping ransomware attacks on user data. In *Distributed Computing Systems (ICDCS), 2016 IEEE 36th International Conference on*, pages 303–312. IEEE, 2016.

[36] Victor J Strecher and Irwin M Rosenstock. The health belief model. *Cambridge handbook of psychology, health and medicine*, pages 113–117, 1997.

[37] Paul Van Schaik, Debora Jeske, Joseph Onibokun, Lynne Coventry, Jurjen Jansen, and Petko Kusev. Risk perceptions of cyber-security and precautionary behaviour. *Computers in Human Behavior*, 75:547–559, 2017.

[38] Ting-Fang Yen, Victor Heorhiadi, Alina Oprea, Michael K Reiter, and Ari Juels. An epidemiological study of malware encounters in a large enterprise. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, pages 1117–1130. ACM, 2014.

[39] Pavol Zavarsky, Dale Lindskog, et al. Experimental analysis of ransomware on windows and android platforms: Evolution and characterization. *Procedia Computer Science*, 94:465–472, 2016.

# A Survey questions

## A.1 Demographics and device details

1. What is your age? [*free text*]
2. What is your gender?
   - Male
   - Female
3. What is the highest level of education you have completed?
   - Some high school
   - High school graduate
   - Vocational training
   - Some college
   - College graduate
   - Some post-graduate work
   - Post graduate degree
4. What is your current employment status?
   - Employed full time
   - Employed part time
   - Unemployed looking for work
   - Unemployed not looking for work
   - Retired
   - Student
   - Disabled
5. What is your race or ethnicity?
   - White
   - Black or African American
   - American Indian or Alaska Native
   - Asian
   - Native Hawaiian or Pacific Islander
   - Other
6. What is your annual household income?
   - Less than $10,000
   - $10,000 - $19,999
   - $20,000 - $29,999
   - $30,000 - $39,999
   - $40,000 - $49,999
   - $50,000 - $59,999
   - $60,000 - $69,999
   - $70,000 - $79,999
   - $80,000 - $89,999
   - $90,000 - $99,999
   - $100,000 - $149,999
   - More than $150,000
7. What is your 5-digit zip code? [*free text*]
8. What is your field of work or study? [*drop down menu*]. Available choices included: Architecture, Engineering, and Math; Arts and Design; Building and Grounds Cleaning; Business and Financial; Community and Social Service; Computer and Information Technology; Construction and Extraction; Education, Training, and Library; Entertainment and Sports; Farming, Fishing, and Forestry; Food Preparation and Serving; Healthcare; Installation, Maintenance, and Repair; Legal; Life, Physical, and Social Science; Management; Media and Communication; Military and Protective Service; Office and Administrative Support; Personal Care and Service; Production; Sales; Transportation and Material Moving;
9. Are you currently using your personal computer (i.e., not one owned by an employer) to fill out this survey?
   - Yes
   - No
10. What operating system do you have installed on your personal computer?
    - Windows

- Mac OS
- Linux
- Chrome OS
- Other

11. What web browser do you typically use on your personal computer?

    - Google Chrome
    - Microsoft Internet Explorer
    - Firefox
    - Microsoft Edge
    - Safari
    - Opera
    - Other

## A.2 Establishing whether a ransomware attack occurred

1. There are many malware attacks that attempt to extort (obtain) money from users. They can be broadly classified into two categories:

   (a) Misleading applications (e.g., fake antivirus scams, spyware removal tools, or PC cleaning apps)

   (b) Ransomware

   Please answer the following questions to help us understand whether you've experienced either of these online scams on your personal computer.

2. Misleading applications usually alert the user to a security issue or vulnerability on their computer, and prompt them to act (e.g., call a tech support number, download or purchase anti-virus software) in order to resolve the issue. Have you ever experienced any of the following scenarios that you suspect were scams? Please select all statements that apply.

    - A security alert or warning popped-up, prompting you to call a tech support number.
    - A security alert or warning popped-up, prompting you to purchase or download software.
    - I have experienced both the above scenarios.
    - I have not experienced any of the above scenarios.
    - I am not sure.

3. [ *Screenshot shown to respondents here (Figure 1)* ] Ransomware is another type of malware that will either lock your computer screen or encrypt your files. If you've been infected with ransomware, you will see screens like the the examples below, informing you that you must pay a ransom to re-gain access to your computer and/or files, providing instructions on how to do so.

4. Have you ever seen a screen similar to the examples above that lock your computer or encrypt your data and ask for money to restore it to normal? Note: These screens are typical of ransomware attacks and will explicitly inform you that your computer has been locked or the files on your computer have been encrypted. It will not tell you to download anti-virus software.

    - Yes, I have seen a screen notifying me that my computer is locked or my data encrypted.
    - No, I have never seen a screen notifying me that my computer is locked or my data encrypted.

5. Some ransomware includes a time limit (typically in the form of a timer counting down), indicating that if you don't pay before the specified time limit expires, then the decryption key will be deleted and your files will be lost forever, or the ransom amount will increase. Have you ever been told that you must pay within some time limit or seen such a timer counting down?

    - Yes, I've seen messages with time limits or timers counting down, telling me I must pay before they expire.
    - No, I've never seen messages with time limits or timers counting down.

6. Some variants of ransomware will encrypt your files so that you can no longer access them. In this case, you might see: (1) Files in all directories with names such as HOW_TO_DECRYPT_FILES.TXT or DECRYPT_INSTRUCTIONS.HTML. (2) Files in all directories with strange extensions such as ".locky". Have you ever had your files encrypted such that you couldn't access them?

    - Yes, I've experienced (1) or (2), or a similar message informing me that my files are encrypted.
    - No, I have never had my files encrypted such that I couldn't access them.

7. **Please read and answer this question carefully!** Have you ever experienced a ransomware attack that informed you your computer was locked or your data encrypted, and asked for money to re-gain access to your computer or files?

    - No, I have not experienced a ransomware attack on my personal computer.
    - Yes, I have experienced a ransomware attack on my personal computer.
    - I am not sure.

8. [ *logic: shown if "I am not sure" selected in Q7* ]. Please help us understand why you have selected "I am not sure". Below are some clarifications about ransomware. Ransomware is a type of malware that will either lock your computer, or encrypt your data. Ransomware will inform users that their computers are locked or their data is encrypted, typically with a large pop-up screen that is difficult to close A common trick is to impersonate law-enforcement agencies and claim that the user has broken the law by downloading copyrighted materials such as pirated music or software, or by viewing other illegal digital materials such as pornography Ransomware will demand money to re-gain access to your computer and/or files, and provide instructions on how to pay Ransomware does not tell users to download software (e.g. antivirus software) to fix the issue.

9. [ *logic: shown if "Yes" selected in Q4, Q5, or Q6, and "No" in Q7* ]. You have reported that you have not experienced a ransomware attack, but have experienced at least one scenario that is typical of ransomware attacks. Why

do you think your experience(s) were not ransomware attacks? Below are some clarifications about ransomware. Ransomware is a type of malware that will either lock your computer, or encrypt your data. Ransomware will inform users that their computers are locked or their data is encrypted, typically with a large pop-up screen that is difficult to close. A common trick is to impersonate law-enforcement agencies and claim that the user has broken the law by downloading copyrighted materials such as pirated music or software, or by viewing other illegal digital materials such as pornography. Ransomware will demand money to re-gain access to your computer and/or files, and provide instructions on how to pay. Ransomware does not tell users to download software (e.g. antivirus software) to fix the issue.

10. [ *logic: shown if "I am not sure" selected in Q7* ]. Please confirm whether or not you've ever experienced a ransomware attack. That is – did you ever see a message informing you that your computer is locked or your data is encrypted which was difficult to close, and which demanded money in order to restore access to your computer or files? Please select "yes" if you've experienced a ransomware attack, regardless of whether or not you paid.
    - No, I have not experienced a ransomware attack on my personal computer.
    - Yes, I have experienced a ransomware attack on my personal computer.
    - I am still not sure.
11. [ *logic: shown if "Yes" selected in Q7 or Q10* ]. [*free text*]. Please describe the ransomware attack you experienced. Do you remember what the message / instructions said? What did the screen look like? Was any functionality of your computer disabled? Please give as many details as possible.

## A.3 Ransomware attack details

The following questions were shown to respondents who reported experiencing a ransomware attack (i.e., selected "Yes" in Q7 or Q10 in the previous section).

1. The following questions refer to the ransomware attack you experienced. If you have experienced more than one ransomware attack, please give details about the most frecent attack.
2. When did the ransomware attack occur? If you don't remember exactly, please give an approximate date (ideally your best guess of the month and year). [*free text*]
3. [*free text*] Do you remember the name of the ransomware? If so, please enter it below. Some examples are: "CryptoLocker", "CryptoWall", "Locky", "TeslaCrypt".
4. How were you asked to pay the ransom (i.e., what was the method of payment used in the attack)?
    - Cryptocurrency (e.g., Bitcoin, Litecoin, Zcash)
    - Payment voucher system (e.g., Paysafecard, MoneyPak, UKash, CashU, MoneXy)
    - Wire transfer
    - Send premium-rate text message to attacker's number
    - Credit card
    - Other
    - I don't remember
5. How much ransom (money) was requested? [*free text*]
6. In the question above, in which currency did you enter the ransom amount?
    - U.S. dollars
    - Cryptocurrency (e.g., Bitcoin, Litecoin, Zcash)
    - Other currency
7. Did the ransomware attack you experience include any of the following characteristics? Please select all that apply.
    - I saw a screen or large pop-up telling me that my computer was locked.
    - I saw a screen or large pop-up telling me that my data or files were encrypted.
    - I was told that unless I paid money, I would not be able to access my files, data, or computer.
    - I saw a timer counting down and was told I must pay money before it expired.
    - I saw a notification page, supposedly from a law enforcement agency (e.g., FBI, Department of Justice, etc.), informing me that I was caught doing an illegal or malicious activity online.
    - I did not experience any of the above.
8. How much time were you initially given to pay the ransom (before the timer expired)? For example, 24 hours, 5 days, 7 days, etc.[*free text*]
9. Did you pay the ransom amount requested ?
    (a) Yes, I paid the ransom amount.
    (b) No, I did not pay the ransom amount.
10. Why did you decide to pay or not pay the ransom? Briefly describe the motivating factors that led to your decision. [*free text*]
11. [*logic: shown if "yes" selected in Q9* ] Was access to your data / computer restored after you paid the ransom?
    - Yes, access was restored.
    - No, access was not restored.
    - I am not sure.
12. Did you notify the authorities of the ransomware attack?
    - Yes, I notified the authorities.
    - No, I did not notify the authorities.
13. Did you try any of the following strategies to remove the ransomware and restore access to your computer or files? Please select all that apply.
    - I re-started my computer.

- I tried to change the extension of files back to their original format and open them.
- I restored my computer from a backup.
- I found and ran a tool to remove the ransomware.
- I found and ran a tool to decrypt my files.
- I used some other strategy.
- I don't remember.

14. Were you able to remove the ransomware?
    - Yes, I was able to remove the ransomware without losing any of my data or files.
    - Yes, I was able to remove the ransomware, but lost my data and/or files.
    - No, I was not able to remove the ransomware. I still can't access my computer and/or files.
15. How did you remove the ransomware?
    - I paid the ransom amount.
    - I re-started my computer.
    - I restored my computer from a backup.
    - I found and ran a tool to remove the ransomware or decrpyt my files.
    - I used some other method to remove the ransomware.
    - I am not sure, I did not remove the ransomware myself.
    - I was not able to remove the ransomware or re-gain access to my computer or files.
16. Did you seek help from anyone else to remove the ransomware? Please select all that apply.
    - I sought help / advice from family and or friends.
    - I sought help from co-workers and/or acquaintances.
    - I sought help from a computer store, repair shop, or other paid IT professional etc.
    - I did not seek help from anyone.
17. What other resources did you use to inform yourself of ransomware, figure out how to remove the ransomware, or to help you decide whether or not to pay the ransom? [*free text*]
18. How do you think you were infected with ransomware? [*free text*]
19. Do you think any of the following actions led you to be infected with ransomware? Please select all that apply.
    - I clicked on a malicious link in an email.
    - I downloaded a malicious program.
    - I clicked on a warning or notification that popped up (either by accident, or purposefully, for example, to close it).
    - I clicked on an advertisement while browsing the internet or on social media (either purposefully, or by accident).
    - I was browsing the internet and did not click on anything.
    - Other
20. Did you change any of your online browsing and/or security behavior following the ransomware attack? Please select all that apply.
    - I changed my operating system.
    - I started backing up my data to an external hard drive or remote file storage server.
    - I bought an antivirus / firewall product.
    - I changed configurations on my computer (e.g., enabled "show file extension" feature, disabled Windows Script Host, restricted login access, etc.)
    - I enabled automatic updates to my operating system, browser, antivirus, and other programs (wherever possible).
    - I changed my default browser.
    - I back up my data more regularly to an external hard drive or remote file storage server.
    - I changed or updated my antivirus / firewall product.
    - I am more careful about which web sites I visit, what I download, and what attachments I open.
    - I update my operating system, browser, antivirus, and other programs more often than before.
    - I encrypted my hard drive.
    - How likely do you think you are to experience a ransomware attack in the future? Please enter a number between 0 and 100, where 100 means: "I'm definitely (100% likely) going to experience a ransomware attack in the future." and 0 means: "There is no way (0% chance) I will experience a ransomware attack in the future."
    - Suppose you were to experience a ransomware attack today and the only way of restoring access to the data on your computer was to pay the ransom (say $300). How likely is it that you'd pay the ransom? Please enter a number between 0 and 100, where 100 means: "I would definitely pay the ransom to restore access to my personal computer and files." 0 means: "No way I would pay the ransom, I would prefer to lose all of my data and files."

## A.4 Security habits

Participants were shown the following prompt at the beginning of this section: "Please answer a few questions about your online habits **right before** the ransomware attack occurred."

1. Approximately how much time did you spend on the internet on your personal computer each day, at the time of the ransomware attack?
    - Less than 1 hour
    - Between 1 - 2 hours
    - Between 3 - 5 hours
    - Between 5 - 10 hours
    - More than 10 hours

2. Approximately how many emails did you open per day on **your personal computer**, at the time of the ransomware attack?
    - Less than 5 emails
    - Between 6 - 10 emails
    - Between 11 - 20 emails
    - Between 21 - 50 emails
    - More than 50 emails
3. How frequently did you download files from online torrent sites such as The Pirate Bay, Extratorrent, TorrentZ2, etc., at the time of the ransomware attack?
    - I frequently downloaded files from torrent sites.
    - I occasionally downloaded files from torrent sites.
    - I rarely downloaded files from torrent sites.
    - I never downloaded files from torrent sites.
4. How did you store information on your computer that you didn't want anyone to see, at the time of the ransomware attack? Please select all that apply.
    - My computer was protected with a password.
    - All sensitive data was stored in a password-protected folder.
    - All sensitive data was stored in an obscure folder that is difficult to find.
    - I only hid data if I expected another person to use my computer temporarily.
    - I immediately deleted all data I don't want anyone to see.
    - I had no sensitive data on my computer.
5. Were you in the habit of backing up your personal files to an external hard drive or a cloud-based storage service, at the time of the ransomware attack? Which of the following statements most accurately describes your behaviour at the time?
    - I did not have any of my files backed up at the time of the ransomware attack.
    - I had been backing up my files approximately once a year.
    - I had been backing up my files approximately every couple of months.
    - I had been backing up my files approximately every couple of weeks.
    - I had been backing up my files approximately every day.
6. Was the hard drive on your personal computer encrypted at the time of the ransomware attack?
    - Yes, my hard drive was encrypted.
    - No, my hard drive was not encrypted.
7. Suppose you have entered your login and password on a website site that you use occasionally (e.g. every two weeks). The browser offers you the option to save your credentials so that they can be used for automatic form completion in the future. At the time of the ransomware attack, what would you generally do?
    - I generally would have saved my credentials.
    - I generally would not have saved my credentials.
8. Suppose Flash Player, Adobe reader, or Flash notified you about updates that need to be downloaded and installed. What wold you generally do at the time of the ransomware attack?
    - I would select "Install updates now".
    - I would select "Remind me later".
    - I rarely see any notifications from such software.
    - I never see notifications from such software.
9. Suppose you are creating a new account on a website that you intend to use occasionally (e.g., airline frequent flyer account). How would you have created a password, at the time of the ransomware attack?
    - I had one password for all my accounts.
    - I had several passwords that I rotated when creating new accounts.
    - I had a password template that I would modify for each account.
    - I'd make up an entirely new one, ensuring that it's strong.
10. Did you own a blog or website at the time of the ransomware attack?
    - Yes, I owned a blog or website.
    - No, I did not own a blog or website.
11. Two-step authentication is an extra layer of security involving two steps to log in to an online account: You'll enter your user name and password. A code will be sent to your phone via text, voice call, or a mobile app. Did you use two-step authentication for at least one of your online personal accounts (i.e., not for a work-related account), at the time of the ransomware attack?
    - Yes, I had two-step authentication on at least one of my personal accounts.
    - No, I didn't have two-step authentication on any of my personal accounts.
12. Did you use a desktop or laptop computer at work at the time of the ransomware attack?
    - Yes, I use a computer at work.
    - No, I do not use a computer at work.
    - Not applicable.
13. [*logic: shown if "yes" selected in Q12*] What task(s) did you use a computer at work for? Please select all that apply.
    - Internet or email
    - Word processing or desktop publishing
    - Spreadsheets or databases
    - Calendar or scheduling
    - Graphics or design
    - Programming
    - Other

| Category | Features |
|---|---|
| Demographics | Gender, race, age |
| Socioeconomic (SES) | Highest level of education completed, household income, employment status, marital status, field of work or study, child under 18 in household |
| Computer knowledge | 8 question multiple choice test (developed by the authors) |
| Security habits | Time spent on the computer each day, number of emails opened per day, frequency of downloading files from online torrent sites, data backup habits (on external hard drive or cloud-based storage device), storage strategy for sensitive information on personal computer (e.g., use of password-protected computer or folder), has encrypted hard drive, credential saving habits in browser, software updating habits (e.g., postpone, install immediately, etc.), own a blog or website, use two-factor authentication (if yes, for which services), password creation habits (e.g., use the same password for all sites), use of computer at work (if yes, for which tasks) |
| Software used | Operating system (name and version), most commonly used browser (name and version), list of plugins installed |

Table A1: *Survey features. Software used were collected passively, and the name of operating system and browser currently used was also asked as a survey question.*

## A.5 Computer knowledge quiz

1. Select the bigger amount of data
   (a) One kilobyte
   (b) One megabyte
2. "Net neutrality" refers to:
   (a) The posting of non-partisan content on websites.
   (b) The manner in which Wikipedia editors are instructed to handle new entries on their site.
   (c) Equal treatment of digital content by internet service providers.
   (d) A promise by users of certain websites that they will not contribute non-partisan comments or work.
3. What does the acronym RAM stands for?
   (a) Random access monitoring
   (b) Running access mount
   (c) Random access memory
   (d) Random access mount
4. Which of the following is an example of an I/O device?
   (a) CPU
   (b) Keyboard
   (c) Power supply
   (d) USB port
5. You are authorizing on a banking website (let's say "Money Bank"). Which web address looks safest to you?
   (a) http://MoneyBank.com
   (b) https://Moneybamk.com
   (c) https://MoneyBank.com
   (d) https://MoneyBank.net.com
6. What is a Trojan horse virus?
   (a) Software that replicates itself to spread to other computers.
   (b) Software that records every keystroke made by a computer user.
   (c) Software that is often disguised as legitimate software.
   (d) Software that encodes itself in a different way (using different algorithms and encryption keys) every time it infects a system.
7. 1 byte consists of ...
   (a) 4 bits
   (b) 8 bits
   (c) 16 bits
   (d) 32 bits
8. Data is permanently stored in:
   (a) RAM
   (b) Hard disk
   (c) CPU
   (d) Cache memory

| Category | Raw proportion | Weighted proportion |
|---|---|---|
| Female | 55% | 54% |
| Male | 45% | 46% |
| White | 81% | 75% |
| Black or African American | 8% | 11% |
| Hispanic or Latino | 5% | 8% |
| Asian | 2% | 2% |
| Mixed | 3% | 2% |
| Other | 2% | 2% |
| Age (19 – 30) | 11% | 16% |
| Age (31– 45) | 19% | 24% |
| Age (45 – 60) | 28% | 27% |
| Age (over 60) | 42% | 32% |
| No High school | 1% | 2% |
| High school | 20% | 32% |
| Some college | 22% | 22% |
| College | 39% | 33% |
| Post-graduate | 17% | 12% |
| Full-time | 38% | 40 % |
| Retired | 28% | 22% |
| Part-time | 11% | 10% |
| Permanently disabled | 9% | 8% |
| Student | 4% | 7% |
| Unemployed | 3% | 5% |
| Homemaker | 5% | 5% |
| Temporarily laid off | 1% | 1% |
| Other | 1% | 1% |
| Married | 51% | 49% |
| Never married | 26% | 31% |
| Divorced | 13% | 11% |
| Widowed | 6% | 6% |
| Domestic / civil partnership | 3% | 2% |
| Separated | 1% | 0% |
| Child under 18 in household - yes | 19% | 25% |
| Child under 18 in household - no | 81% | 75% |
| Less than $10,000 | 3% | 4% |
| $10,000 - $29,999 | 8% | 8% |
| $20,000 - $29,999 | 10% | 10% |
| $30,000 - $39,999 | 11% | 12% |
| $40,000 - $49,999 | 9% | 9% |
| $50,000 - $59,999 | 10% | 11% |
| $60,000 - $69,999 | 7% | 6% |
| $70,000 - $79,999 | 8% | 7% |
| $80,000 - $99,999 | 9% | 8% |
| $100,000 - $119,999 | 6% | 6% |
| $120,000 - $149,999 | 5% | 5% |
| $150,000 - $199,999 | 4% | 4% |
| $200,000 - $249,999 | 2% | 1% |
| $250,000 - $349,999 | 1% | 1% |
| Prefer not to say | 8% | 9% |

Table A2: *Demographics and socioeconomic status of respondents, n=1,180. The raw proportion represents the fraction of respondents out of n=1,180 having a particular characteristic, and the weighted proportion represents the post-stratified proportion.*

# Enhancing Privacy through an Interactive On-demand Incremental Information Disclosure Interface: Applying Privacy-by-Design to Record Linkage

Hye-Chung Kum[1], Eric D. Ragan[2], Gurudev Ilangovan[1], Mahin Ramezani[1], Qinbo Li[1], Cason Schmit[1]
[1]*Population Informatics Lab, Texas A&M University* [2]*INDIE Lab, University of Florida*
{kum, ilan50_guru, mahin, lee, schmit}@tamu.edu; eragan@ufl.edu

## Abstract

Achieving the benefits of data science in cases involving personal data requires the use of that data, which results in some privacy risk. Our research investigates approaches to enhance privacy while supporting legitimate access for human decision making by capitalizing on the fact that in most human-computer hybrid systems, only a small fraction of the full data is required for human judgment. We present an interactive visual system for record linkage – a task that requires human decision-making about whether different but similar data records refer to the same person. The system employs an on-demand interactive interface that incrementally discloses partial information only when needed and other feedback mechanisms to promote ethical behavior. We evaluate our approach with a controlled experiment of how different types of feedback and access restrictions affect human decision-making quality, speed, and access behavior. The on-demand interactive interface reduced privacy risk to only 7.85%, compared to 100% when all data is disclosed, with little to no impact on decision quality or completion time. In addition, feedback from an expert review supports the notion that an intermediate level of access other than "all or nothing" can provide better accuracy than no access but more protection than full access.

## 1. Introduction

The potential impact of population informatics—data intensive secondary analysis of large integrated population data—are endless [1]. Access to such data for qualified researchers could provide a greater understanding of root causes of social and public health problems, help identify upstream opportunities for interventions, help predict the downstream effects of different policy options, and assist in allocating our collective resources for the greatest impact to benefit our society. As one example, a National Institute on Drug Abuse (NIDA) study integrated data from multiple databases on 56,923 Medicaid beneficiaries with opioid dependency to conclude that buprenorphine was cheaper and safer than alternative treatments [2, 3]. Another example is a three state study that integrated three data systems to follow children from the foster care system for over 10 years to investigate long-term employment and income trends [4].

| ID | First name | Last name | DoB(M/D/Y) | Sex | |
|---|---|---|---|---|---|
| 8002767 | JUDE | WILLIAM | 09/09/1906 | M | Maybe Father-Son |
| 8003567 | JUDE | WILLIAM JR | 09/09/1960 | M | |
| 0006947 | BRYANT | MADELINE | 05/02/1962 | F | Probable Data error |
| 0006947 | MADELINE | BRYANT | 05/02/1962 | F | |
| 9018540 | SALLY | BYRD | 07/04/1960 | F | Maybe Twins |
| 6008928 | JOHN | BYRD | 04/07/1960 | M | |

Figure 1: Pairs of PII for human judgment in record linkage

While the potential benefits of population informatics are clear, access to such population data for these research is not widespread and is often given on a one-time basis for a single project. In fact, when compared to the widespread use of population data in other less regulated sectors such as marketing, intelligence, and campaigning, secondary analysis of population data for research is quite restricted and lacks infrastructure. This is especially true in the United States, where privacy concerns make it difficult to build and maintain large integrated population data for research. In contrast, Canada, UK, and Australia have invested in establishing population data linkage centers [5].

One of the core challenges in establishing integrated population databases is addressing privacy concerns during data integration. High-quality data integration requires record linkage (RL), the process of identifying records from heterogeneous data sources that potentially refer to the same person in cases where a common identifier is not available. Privacy becomes a major issue because one must exactly identify the identity of records to accurately build the integrated data. During RL, it is important to distinguish between family members or twins [6] and to handle changes in the data (e.g., change of last name) and data errors. Thus, most projects that require integrated data obtain access to *personally identifiable information* (PII) (e.g., names, birth dates) to use for RL. Figure 1 shows a simple example of different types of data discrepancies in PII pairs.

In practice, most linkage projects use semi-automated linkage systems where the majority of the linkages are made using algorithms that humans have to tune, maintain, and manually resolve more complex cases. Properly using auto-



*USENIX Symposium on Usable Privacy and Security (SOUPS) 2019.* August 11 -- 13, 2019, Santa Clara, CA, USA.

mated methods requires a significant level of human involvement for data cleaning [47], standardization, training data construction, parameter tuning [46], and validation [6, 7]. For example, Bronstein et al. [8] describe the process of matching pregnancies from Medicaid data to birth records using probabilistic record linkage that involved 11 manual steps. The process required many human decisions in the process such as cleaning up 4,369 pregnancies linked to more than one vital records. In another study linking cancer registry data to health service data, 15% (16,288 links) were confirmed through manual verification [9]. Such a high level of human interaction and iteration is common in medical record linkage studies [8-10]. Without these human involvement, the match rate and biases in these studies would be problematic because the errors in the linkage step propagate downstream to all analysis using the integrated data. In addition, automated linkage can also result in selection bias such as in preferentially selecting patients with complete information on required identifiers, which can under-represent particular groups such as socioeconomically disadvantaged and racial/ethnic minorities [5, 8, 11, 12, 13].

Human review often involves looking at similar pairs of records to make judgments about complex cases where different records could potentially refer to the same person or to decide how to tune linkage algorithms to improve data integration. These tasks require disclosure of some PII to some people, which poses potential privacy risks. In most linkage projects, this usually means that staff get full access to all PII in a dataset to enable quality linkage decision-making [5, 8, 9]. Naturally, such unlimited access to PII raises concerns about privacy and has led to many efforts to develop automated privacy preserving RL algorithms. Most algorithms securely compute a known linkage function using encryption and trusted third-party computing. These approaches assume *no access to PII as a solution to privacy*. However, such solutions are problematic for assuring quality data integration for complex cases requiring human judgment, especially since the details of the linkage function is rarely known ahead of time. Thus, open questions remain about (1) how to determine the linkage function, (2) support the human tasks required to obtain high-quality linkages, and (3) how to validate the linkages found [14, 15].

Our work on privacy-enhanced RL takes a fundamentally different approach to privacy in order *to produce high-quality validated results by enabling human judgment where needed.* Rather than rely on various security technology to limit access for privacy, our premise is that human access to some PII is necessary and encouraged for valid results. Hence, we rely on two fundamental principles of privacy to promote legitimate and confidential access: (1) *the minimum necessary* principle and (2) *accountability through transparency* principle. Thus, the critical questions for privacy enhanced system design are: (1) What and how much information about the PII needs to be accessed for good linkage decisions? (2) When do you know what you need to access? (3) What accountability mechanisms would be effective to discourage bad behavior? Our approach is motivated by the hypothesis that there is some level of partial disclosure of PII—between unrestricted access to all PII and no access to PII—that can effectively support human judgment and validation while significantly reducing total PII access.

In this paper, we present and evaluate a novel privacy-enhanced RL system (see Figure 2) for safe human interaction with PII. The core tenet of our method is an on-demand interactive method for incrementally disclosing limited information, only as-needed, and when explicitly requested. This approach makes it possible to meet the legal requirements for minimum necessary information disclosure standards while also enabling accountability through the ability to log access requests to individual PII details. The contributions of the presented research are threefold:

- First, we present our interactive record-linkage system that uses (1) on-demand, incremental information disclosure, (2) feedback of privacy risk, and (3) enforcement of disclosure budgets to facilitate high-quality decision-making while limiting overall access to personal data.
- Second, we present a controlled experiment to evaluate how different types of feedback and access restrictions affect human decision-making quality, speed, and access behavior in a record linkage task.
- Third, we also present an expert review with domain scientists who regularly conduct research with PII.

## 2. Background and Related Work

In this section, we provide the basis of our approach in the privacy regulations and review relevant privacy literature.

### 2.1. Minimum Necessary Standard and Practical Challenges

As discussed in detail in the introduction, human review of personal data is common for a variety of data work and required for valid results [6-10, 46, 47]. Research on information privacy has shown the complex nature of providing protection while still allowing utility from legitimate use of personal data for social benefit [16]. Among the core principles for designing privacy-enhanced systems is to limit disclosures of protected information to only those necessary for achieving a given purpose. This principle is central to many different data protection laws in the form of *minimum necessary* or *need-to-know* information disclosure standards. Laws like the *Health Insurance Portability and Accountability Act* (HIPAA), the *Privacy Act of 1974*, and the confidentiality protections for substance abuse disorder records in *42 CFR Part 2* use similar legal standards to permit legitimate uses of data while protecting privacy by limiting extraneous disclosures [50-52]. Similarly, the EU General Data Protection Regulation (GDPR), uses the principle of "data

minimisation" to limit data use to what is necessary for a permitted purpose [53].

However, practically implementing a process for sharing protected data that restricts disclosures to the minimum necessary is a daunting task [54]. It is rarely the case that a data project knows exactly what data elements and observations are needed ahead of time. Instead, data science is often an iterative process of learning from the data and refining the analysis until useful results are obtained. Moreover, the iterative nature of analytic methods also means that the required data dynamically changes over the course of the project. Practically, in many situations, it is the case that *all* the data is decided to be the "minimum necessary" [55].

These dynamics can lead to serious consequences when negotiations and legal agreements must be made (e.g., data use agreements) between different organizations for data sharing. Perceptions about what constitutes the minimum necessary can differ between data sharing partners, leading to prolonged project delays [56]. Even worse, funded projects may be cancelled when researchers are not able to pass a vetting process for giving full access to protected data [56]. One reason for this is because there are no practical tools to facilitate data disclosures that better meet minimum necessary legal standards. Our research addresses this need.

### 2.2. Privacy and Human Behavior

Researchers have explored a variety of approaches to system design [17] and interface design to support privacy enhancements [18]. For example, Iachello et al. [17] describe the design process for a privacy aware social location disclosure application through a series of user studies. They present a list of privacy guidelines from these studies demonstrating the privacy by design approach. Dasgupta et al. [19] presents metrics for privacy as applied to visualization. They demonstrate the use of aggregation, clustering, and uncertainty in scatterplots and parallel coordinate plots to allow inspection of sensitive data while limiting knowledge of individual elements. In contrast, our work focuses on data inspection tasks that require review of individual PII for accurate decision making.

As an example from our prior work involving access to individual PII, Ragan et al. [20] demonstrated how the use of visual masking techniques could be used to hide data values in tabular data interface while still showing differences to support data cleaning and de-duplication tasks. Kum et al. [6] also studied different mechanisms (i.e., deception, obfuscation, and blurring about the nature of the list of names) to hinder inference of identity when names are disclosed. They found that these methods were effective in introducing uncertainty to protect the real identities of names for both common and rare names. Work by Hasan et al. [21] used a similar approach but for images. The authors studied visual obfuscation methods for hiding or altering portions of photographs to preserve privacy, and they discuss the tradeoffs of different approaches in terms of both privacy and the effects on the general interference or distraction when viewing images. Similarly, Çiftçi et al. [22] demonstrated how altering the color composition for facial images can make it difficult to recognize people in photographs.

Prior research has also demonstrated that users' behavior or attention to privacy can be influenced by their experiences with technical systems. For example, Chang et al. [23] found that participants' inclination to disclose information could be influenced by the types of profile pictures they observed prior to the interaction. When viewing less revealing profile images, the participants were less likely to share their own personal information. These results suggest that decisions for acceptable privacy behavior might be influenced by the perception of what others find acceptable. John et al. [24] ran similar experiments asking participants whether they had engaged in a number of sensitive activity (e.g., sexual behaviors). They measured the proportion of questions answered affirmatively as an indicator for privacy concerns and varied the look and feel of the website (i.e., professional, baseline, unprofessional). Those who were asked on the unprofessional website were almost twice as likely to admit to engaging in the sensitive activities compared to the baseline and professional websites indicating that disclosure of private information responds to environmental cues. The results support the general idea that a system may influence a user's attention to privacy by using different cues.

In fact, a comprehensive review of multi-disciplinary literature presents multiple interventions (e.g., education, feedback, framing, positive and negative incentives) that can be used to influence privacy decision making [25]. In this research, the main intervention tested is feedback. Password meters is a good example of how effective feedback systems can nudge to create stronger passwords [26]. Our research studies a feedback mechanism similar to the password meters that gives real-time feedback and allows decisions to be altered based on the feedback given.

### 2.3. Quantifying Information Privacy

Quantifying the privacy risk is an active area of research with the best approach being context dependent [27]. k-anonymity was the first method proposed based on the insight that a record may not be distinguished from at least k-1 records when there are k shared records [28-31]. Machanavajjhala et. al. [32] have shown issues with k-anonymity when there is a lack of diversity allowing for background attacks and introduced the l-diversity model that aims to have intra-group diversity for sensitive values. Li et al. [33] have shown that (1) l-diversity may be difficult and unnecessary to achieve and (2) l-diversity is insufficient to prevent attribute disclosure. To address these problems Li presented

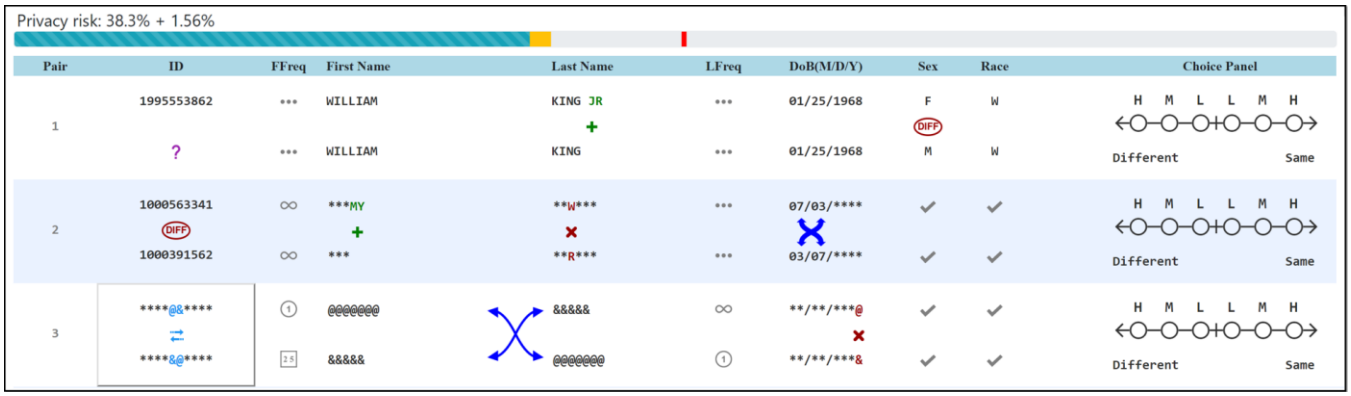



Figure 2: Example from the study application showing (1) supplemental markup and value masking, (2) interactive clickable interface, (3) feedback privacy meter, and (4) the privacy budget (solid red line on the meter). The visual markup highlights discrepancies, provides information about name frequency, and hides common values. The box on the last row indicates that the user has moused over this area, and is considering whether to click open or not. The user should be taking into account the feedback meter on top which indicates the accumulative disclosures to now in blue, and what additional risk will occur if the selected information is clicked open in orange. Finally, the solid red line on the meter indicates a limit to the disclosure that the user can request.

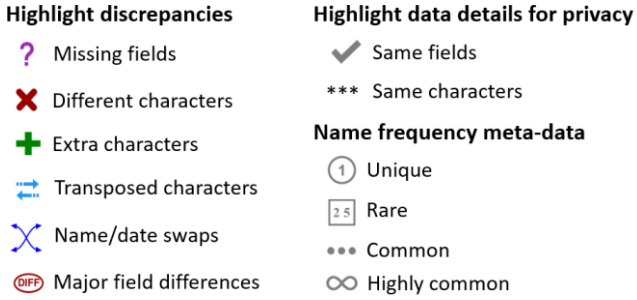



Figure 3: Visual masking icons used to highlight discrepancies, including matching values, and providing meta-data [30].

t-closeness based on the differences in the distribution of the sensitive attribute [33]. In another study, Li et al. [34] presented another approach based on k-anonymity and differential privacy and used a method of input perturbation to add uncertainty. Currently, differential privacy models provide the strongest guarantees and is the most active area of research [27, 35]. In particular, many differential privacy algorithms have been proposed to answer low dimensional counting queries. However, adoption of these methods in practice has been limited due in part to the wide variation in error rates, which are dependent on the properties of the input data [27]. It is important to note that although these approaches are related and may be applicable to the work in this paper, quantifying the identification risk to support user decisions to disclose certain parts of the PII in RL is different from risk in low dimensional counting queries. Although k-anonymity does not address sensitive attribute disclosure, it is a well-established measure for identity disclosure, the focus of this work [36], and our paper presents our first approach based on k-anonymity with the incorporation of differential privacy in progress.

## 3. System Design

Our research contributes a novel interactive interface where we start with fully-masked de-identified data and let users click to open when more information is required for good decisions. The interface is meant to serve as a complement to algorithmic methods [15] for detecting possible duplicates or discrepancies among similar records. For uncertain cases requiring human review and judgment, the system presents the flagged pairs as rows in a tabular interface with different data fields separated by columns (see Figure 2). The system takes advantage of three techniques to enhance privacy protection: (1) minimum necessary disclosure via just-in-time, incremental information access, (2) transparent accountability by quantifying the privacy risk due to the disclosure made, and (3) limiting data access via a budget. Before we present our study of how these techniques can affect privacy while still maintaining the quality of the linkages made, this section describes the system in terms of its core mechanisms and design rationale.

### 3.1. Design Rationale

Our research capitalizes on the fact that in most human-computer hybrid systems for sensitive data, only a tiny fraction of the full data is required for tasks requiring human judgement. Prior research provides evidence to suggest that the optimal level of disclosure to achieve high quality linkage is quite low with minimal risk to identification when appropriate meta-data is shared using data masks [20]. Ragan et al. investigated the effectiveness of different levels of disclosure on static interfaces [20]. While the results are promising, the tested system only supported static, pre-specified levels of data hiding.

Our research investigates dynamic *just-in-time incremental* techniques to enhance privacy in data systems requiring human access to personal data for legitimate purposes. We present an interactive visual system (see Figure 2) for linking personal data using an on-demand interface that incrementally discloses limited information—and only when

needed and explicitly requested. This approach minimizes data disclosure to optimal levels for human judgment while observing legal requirements to follow a *minimum necessary disclosure* standard. Further, the system's interactive interface satisfies accountability requirements since all disclosure occurs via explicit user actions, which makes it trivial to log who accessed what data.

In addition, we study different design mechanisms to promote accountable ethical behavior in information access decisions. The system incorporates visual feedback and allows access limitations to encourage conscientious data review. By quantifying and displaying the privacy risk associated with each increment of disclosure, the system can help users to consider the tradeoffs between privacy and decision quality for each piece of information. Further, an optional maximum *disclosure budget* can be enforced to provide guidance to novice users about the right balancing point for good decision making, or to meet external requirements for accessing sensitive data.

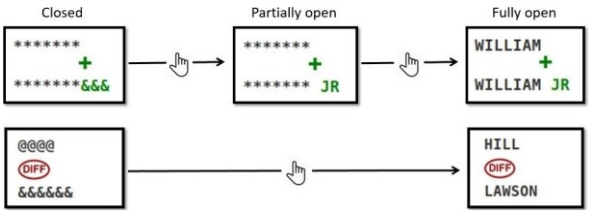



Figure 4: Interactive on-demand interface. Cells start with no disclosure, then partially open with a click. Cells open fully with either 1 or 2 clicks depending on the nature of the data.

### 3.2. Minimum Disclosure via Interactive Just-in-Time Interface

The system is designed to support minimum disclosure for interactive RL using dynamic information access. When linking records, the reviewer's task is to consider the discrepancies and make a decision whether the two records corresponds to the same person or different people. A choice panel on the right side allows fast recording of each decision (shown on the right-most side of Figure 2), and the interface can also record the degree of user confidence by high, moderate, and low levels (denoted by *H*, *M*, and *L* labels).

The first part of enabling the on-demand design is to have the default state of the interface hiding all characters and instead use visual icons and meta-data to help highlight how the records in each pair differ. Figure 3 shows an overview of the primary icon types used for visual masking [20]. The bottom row of Figure 2 (*Pair #3*) shows an example of this default masked state. As seen in the "Sex" and "Race" fields, checkmark icons indicate cases where the field contents are exactly the same. Otherwise, asterisks and replacement characters summarize differences, and supplemental icons help explain the type of differences. Additionally, icons indicate frequency of both the first name and the last name (by the *FFreq* and *LFreq* columns, respectively) in each data source since linkage decision-making can depend on how rare or common names are in the source data. A more detailed explanation of the visual masking techniques and an evaluation of their effectiveness for decision making are presented in [20].

The focus of this paper is the creation and evaluation of a new method for accessing details only as needed. As seen in Figure 4, the user is first presented with only the de-identified data using visual masks. If users need more information to make good linkage decisions, they can click on the cell (i.e., a specific field for each PII pair) to reveal more details. The first click will disclose only the characters that are different between the cells for that pair, and an additional click on that cell will show the full cell contents. Depending on the nature of the differences in the pair, cells will open fully with one or two clicks. While users have the choice to access different levels of detail, the example pair in the middle row of Figure 2 (*Pair #2*) shows examples of partial disclosure, where some characters are visible to aid inspection of differences. The top row (*Pair #1*) shows full disclosure without character masks, which would be possible if the user fully clicked each individual cell in the row. By enabling different levels of disclosure that only reveal data with an explicit click, the system effectively supports the *minimum necessary* principle by preventing most of the PII from being accessed unless necessary for good decisions.

### 3.3. Accountability via Quantified Privacy Risk

While the interactive method for just-in-time incremental disclosure can support minimum necessary access to PII, it allows users to decide what information is accessed. To promote preferred privacy-aware behaviors and prevent misuse of sensitive data, the interactive interface is augmented with mechanisms for information accountability (i.e., transparency and continued monitoring) and audits for misuse [37]. Concretely quantifying the privacy risk and making this information available to everyone (e.g., data workers, managers, and compliance administration) is the start of accountable access to PII.

Quantifying the privacy risk and providing feedback on it also supports good human decisions because decisions can be improved when it is informed by relevant information [26]. To impact the decision, the information must be concrete enough to be actionable. Generally, instructing researchers working on linkage to "disclose as little as possible" is not actionable. But, when the actual risk of identification from disclosing a piece of data can be concretely quantified and shared ahead of time to inform the decisions to view the data, we believe that people will be encouraged to make more thoughtful decisions based on the risk level.

To this end, our system uses a method for quantifying privacy risk based on factors such as amount of characters accessed, type of information, and its uniqueness; then, the interface uses this privacy measure to display feedback

about the risk associated with each data-access decision. Feedback is shown by a visual meter (see the top of Figure 2), where the length of the meter represents 100% disclosure, the blue bar represents the current accumulative access, and the temporary orange extension represents the added risk of disclosure for the currently selected cell that the user has moused over. If the user decides to click on a masked cell, then the data will be shown, the privacy cost will be used, and the meter will update to show a new level of disclosure. If the user moves the mouse off of the cell without clicking, the "hypothetical" orange increase to the meter bar goes away. The recording of these clicks and the feedback have similar role to how a surveillance camera can encourage good behavior.

A number of factors were considered for risk quantification. Measuring the identity disclosure risk for a given partial disclosure of personal data is not trivial because not all pieces of information lead to the same level of identification. Mathematically, the identity disclosure risk is inversely related to the number of entities in the population that share the information disclosed. If the information refers to one and only one person in the population, then the uniqueness of a person's identity information makes it easy to match the information to a specific person. On the other hand, if the disclosed information is identical for multiple people, then the information is less revealing, as it could refer to any one of those people. Quantifying privacy risk is an active area of research with the best approach being context dependent [27].

For our system, the goal is to quantify the identity disclosure risk because sensitive attribute disclosure is fundamentally blocked by keeping the sensitive attributes separate from the identifiers. Thus, our prototype used the k-Anonymity Privacy Risk (KAPR) score which uses the anonymity-set size as an estimate of the identity disclosure risk. *Anonymity-set size*, defined as the number of people in the population who share the same identifying information, is an intuitive and accessible measure to estimate the privacy risk. The larger the set size, the lower the privacy risk. For example, when a frequently occurring name (e.g., Eric) is disclosed, there is a low probability that a specific person with that name could be identified. In comparison, a rare name (e.g., Mahin) may be sufficient information to determine a person's identity. In addition, anonymity-set size is easily calculated dynamically for any information to be disclosed during human interaction with the system. As more information is disclosed to aid linkage, the anonymity-set size will be reduced. This in turn will increase the privacy risk.

In sum, The KAPR score is a normalized score from 0% (nothing disclosed) to 100% (everything disclosed) with higher scores if more is disclosed and what is disclosed is more unique. Uniqueness is calculated based on the data being linked. An example and the exact measure can be found in [38]. Although the KAPR score function was used in our meter in the user study because of its accuracy for measuring identity disclosure, it is important to note that the exact function used is not as important as the use of a reasonable feedback meter that users can understand. Further research is needed to study the trade off between using easy to understand functions (e.g., percentage of information disclosed) versus more accurate but complex function (e.g., KAPR score) for quantifying the privacy risk.

#### 3.4. Limiting Privacy Risk via Budget

Although the interactive interface enables only the minimum necessary disclosure and the feedback meter encourages limited access behavior through accountable access to PII and audits after the fact, neither of these designs alone can enforce limited disclosure that may be a condition of use. For example, certain data usage agreements may limit access to social security numbers by allowing up to four digits. In our system, such hard rules on data access can be enforced using an option to configure the interactive interface with hard rules ahead of time. In particular, the privacy budget feature can be used to enforce a limit on the total disclosure for a given use case.

By specifying an allowable privacy budget ahead of time, the system can guarantee a certain level of information disclosure. Moreover, specifying a budget based on expert users can provide guidance to novice users about the right balancing point between access to data for good decisions versus trying to make do with limited access to information which can result in lower quality decisions.

Ultimately, the goal of any legitimate access to sensitive data is to maximize utility under a fixed privacy budget. Thus, it is important to design the system that allows for specifying the privacy budget ahead of time so that it can be enforced. Figuring out appropriate levels of privacy risk for a given task to support quality data is an open research area that will require further research. In our evaluation, we start by studying how different privacy limits might lead to different human behavior in making decisions to disclose information, as well as how these limits on the privacy score impact the quality of the record linkage task.

### 3.5. Threat model

The main threat model for this work is the insider threat model where the system goals are to minimize any incidental knowledge from legitimate access to PII, and discourage against access for unauthorized purposes by authorized users. First, the on-demand interface will minimize any incidental privacy risk of data workers seeing information about people they know (e.g., co-workers). In addition, quantifying the privacy risk with the meter feedback discourages insiders from abusing their ability to access information. This is similar to surveillance cameras that

discourage people from bad behaviour by making it possible to enforce accountability. To operate the system effectively, having clear reporting and audit processes in place for the logs will be important just as with camera footage. Although cameras cannot gurantee no bad actors, it is very effective in keeping people on good behaviour, especially when it is clearly displayed. Finally, enforcing a limited budget provides further ability for mangers to manage risk from insiders at acceptable levels. Managers may set low limits on disclosure ahead of time and iteratively increase the limit as requested when the context requires high levels of privacy.

# 4. Experiment

Using our interactive record-linkage system, we conducted a controlled experiment to evaluate how different mechanisms for privacy protection affect information access and decision-making for tasks requiring interpretation of PII.

### 4.1. Hypotheses

Our over-arching goal is to design and evaluate effective ways to discourage unnecessary information disclosure without increasing linkage errors. In this experiment, we test the effect of the following three mechanisms, 1) an interactive clickable on-demand disclosure interface, 2) transparent accountability through measuring the real-time risk on a meter, and 3) enforcing limitations on disclosures through a pre-specified budget on the meter. Our evaluation of these mechanisms follow three respective hypotheses:

**H1**: We hypothesize that an appropriate on-demand and incremental disclosure interface can significantly reduce disclosure without compromising decision quality. This is the main premise behind our design for interactive on-demand information access. An explicit click by the user is required to disclose any piece of PII which means that all clicks, and thus disclosures, can be tracked. Given that users will have the ability to look at any part of the PII, there should be no impact on the quality of the decision.

**H2**: The second hypothesis is that the addition of the feedback mechanism, which quantifies and provides a real-time display of consequences of the click, can better inform the decision to access information, and hence encourage only the most needed disclosure. The quantification of the risk and visibility of this information for all relevant parties (e.g., users, managers, compliance) will discourage misuse of PII and encourage accountable use of PII through transparency.

**H3**: The third hypothesis is that when providing feedback on disclosure, enforcing a limit on privacy disclosure through a pre-specified budget will change disclosing behavior to tend toward the given limit. That is, we expect people will naturally try to use the full available budget. In other words, if the limit is set high, then higher levels of disclosure will occur (**H3.1**). On the other hand, if the limit is set too low, disclosure levels will be forced to be lower, but decision quality will be negatively affected (**H3.2**). Hypothesis **H3.2** follows the results in [20], which provided evidence of a limit to how much data can be hidden before negatively influencing the quality of judgment in decisions involving person-level data.

### 4.2. Experiment Design

To address our hypotheses, the experiment followed a between-subjects design with the following five conditions:

- *Fully open*: non-clickable interface with all details already visible: This was the baseline condition used to study the effect of different mechanisms. It used the static full disclosure interface with visual discrepancy highlighting and frequency meta-data, but no data was hidden.
- *No meter*: clickable on-demand disclosure with no feedback meter, and no limit. The goal for this condition was to test the effect of using an interactive on-demand interface on the amount of disclosure and decision quality. The initial interface starts with a fully-masked display with markups, and users can click to disclose more information. The KAPR feedback meter was not shown, and there was no limit to information access.
- *Unlimited meter*: clickable on-demand disclosure with an unlimited feedback meter. The goal of this condition was to test the effect of adding the KAPR meter (see top of Figure 2) to display the potential real time increase in risk for any given disclosure to inform the decision to view the data. There was no limit to disclosure in this condition.
- *High limit*: clickable on-demand disclosure with a feedback meter and a high limit. This condition tests the effect of enforcing a pre-specified limit on the privacy budget indicated by a thick red line on the meter. This condition sets the limit at a moderate disclosure level believed to be sufficient to make good linkage decisions. The specific limit in this condition was 35.7% to 37.8% KAPR score depending on the specific dataset. This amount was chosen based on the *moderate* level from a prior study [20] that focused on static, non-interactive levels of information disclosure. The prior study found this level of disclosure had comparable decisions as full disclosure, so we would expect good linkage performance if participants used the full budget.
- *Low limit*: clickable on-demand disclosure with a feedback meter and a low limit. This condition is similar to the previous condition in enforcing a limit on the privacy budget as in Figure 2. This condition sets a lower limit with KAPR scores ranging from 5.02% to 6.48% depending on the dataset. This level was again chosen based on a previous study [20], which found reductions in linkage decisions with this amount of static disclosure. In the current study, users choose which details to access interactively, as needed. Thus, this condition tests whether total disclosure levels can come down to these low levels with-

out compromising linkage decisions when interactive disclosure is used.

Figure 5 shows a simplified summary of the differences among the five conditions. The conditions allow us to test our hypotheses about the effects of different mechanisms to discourage unnecessary disclosure. We address hypothesis H1 by comparing the results from the *fully open* to the *no meter* to determine how much more we can reduce disclosure using the interactive interface. Hypothesis H2 compares the *no meter* to the *unlimited meter* to determine if a feedback meter is effective in reducing unnecessary disclosure. Finally, hypothesis H3 compares the *unlimited meter*, *high limit*, and *low limit* to evaluate the impact of different levels of limit on the amount of disclosure and quality of linkage.

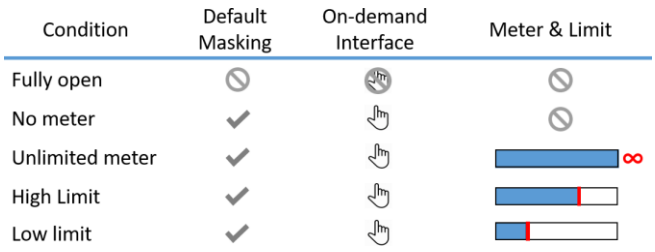



Figure 5: Visual summary representing the differences of the five experimental conditions in the evaluation.

### 4.3. Generation of Test Data

To allow us to evaluate the effects of the different system configurations on record linkage performance, we had to have data pairs that could serve as "ground truth". Since real scenarios do not have known "true" answers, our experiment used a derived data set created by modifying publicly available voter registry data (as in a previous study [20]). The generated test data comprised of realistic pairs of records based on a large county's records from 2013 and 2017. To establish a known ground truth, the registry number and address information were used to identify the same people among many generated pairs in the original data. We also tweaked the pairs to control for the kinds of differences and emulated real world data errors like typographical errors, family relationships (e.g., twins and siblings), name changes, field swaps, and missing fields.

In total, we had 747 pairs of records with "same" or "different" labels. These pairs were used to generate 10 random samples of 36 questions each, and each user was randomly assigned one such sample. It should also be noted that out of the 36 questions, 6 questions (one in each page) were easy questions for which the answers were obvious (for example, all different fields would mean the pair referred to different people and vice versa). These questions primarily served as attention checks and to verify that participants had sufficient understanding of the decision-making process for linkage.

### 4.4. Procedure

The study was approved by our organization's Institutional Review Board. We note that the procedure for this study was designed to be similar to a previous study using an interactive record linkage activity [20]. The study was run in group sessions in a computer lab, but each participant worked independently. Each study session lasted two hours. The system was run as a web application on Google Chrome. All participants used identical computers running Windows 7 with 23-inch displays at 1920x1080 resolution.

To begin, participants completed a background questionnaire to collect information about age, gender, education, academic specialization, experience with data analysis, and primary language. Next, the experimenter gave participants an overview of record linkage, the system, and the instructions for the task. Participants then worked through the system's tutorial, which included sample questions and additional instructions. To help participants understand the decision making, the tutorial provided the correct answers for any practice linkage questions that were answered incorrectly, and participants had the option to repeat or review all information. Different configurations of the tutorial were designed to match each experimental condition, and the final phase of the tutorial had participants work through 36 practice questions.

After the tutorial, participants started the main linkage trials, which were organized into multiple sets of 36 linkage questions shown in groups of six questions per page. Participants worked through as many sets as they could complete in the study time. Finally, near the end of the study session, participants concluded by completing a closing questionnaire that asked for comments about the linkage task, the system, and their comfort with sharing personal information.

The paper presents data from the first set to 36 linkage questions because everyone completed at least one. Analysis of the second set had comparable results with the first set, but fewer participants. Full analysis scripts and data are publicly available at [48].

### 4.5. Participants

The study had a total of 122 participants, and each participant completed one condition. We used the 6 trivially easy questions to filter participants who did not demonstrate sufficient effort or competency for the linkage task. Participants who incorrectly answered more than one of the easy questions were excluded from analysis. Two participants failed to meet the requirement, and hence their data was excluded. Thus, data from 120 participants were considered for data analysis. Of these, 22 were in fully open, 23 were in no meter, 26 were in unlimited meter, and high limit, and 23 were in low limit. The final numbers in each group varied due to the competency screening and the study being run in pre-scheduled lab sessions. 55.8% of the participants were female and 44.2% male. Ages of the participants ranged from

18 to 42 with a median age of 22. Participants came from diverse academic fields. 52.5% of participants were either pursuing a graduate degree or already had one, and the remaining participants were undergraduate students.

## 5. Experiment Results

We analyzed the results to test for differences based on the previously explained hypotheses. We did not conduct statistical comparisons of all five conditions together because this would confound the presence/absence of different mechanisms. When testing for statistical differences in measure (error rate, KAPR scores, duration), we conducted either a parametric test (Student's t-test or Welch's t-test) or a non-parametric test (Wilcoxon rank sum test) depending on the data and if the assumptions of parametric testing were met. The procedure we followed was to test the measure for normality using the Shapiro-Wilk test. If the measure passed the normality assumptions, we did an F-test to test for the homogeneity of variance. Further, we tried data transformations (e.g., log or square root) to satisfy assumptions when possible. If the measure passed the normality and F-tests, we used a Student's t-test. If it passed the normality test but failed the homogeneity of variance test, we used the Welch's t-test which accounts for different variances. If it did not pass both, we used the non-parametric Wilcoxon test. For all tests, we used a base alpha level of 0.05 and applied Bonferroni correction for the four hypotheses, which resulted in an adjusted significance threshold of 0.0125.

### 5.1. Performance Overview

Risk of privacy loss was calculated using the KAPR measure which calculates the actual risk of identification (i.e., how unique the revealed information is) based on what information has been disclosed. Across all conditions, the score ranged from 0% to 100%, with overall mean of 23.31% (SD = 36.79). Figure 6 and Figure 7 show the error rate and risk score results broken down by condition. We present the results using violin plots which, in addition to displaying the median and interquartile range, also show the distribution of the data [57].

We also consider completion time (see Figure 8), which includes only the portions of the study spent answering the main 36 record-linkage questions. Analysis of the differences in participant confidence in linkage decisions had similar results to a previous study where confidence was lower for incorrect responses, which suggests their lack of confidence was justified [20].

### 5.2. H1: Effects of interactive on-demand disclosure

Hypothesis 1 is concerned with differences in information disclosure between the baseline static interface with all information already visible and the on-demand interface starting with no information but incrementally reveals more when participants need to see more. So we compared conditions *fully open* and *no meter*. The *fully open* condition had full data disclosure with no data hiding, so the privacy risk score was a constant 100% KAPR for all participants. In comparison, *no meter* had an average of only KAPR = 7.85% (SD = 5.23), indicating very low levels of disclosure and risk even though the participants could reveal as much information as they wanted. While this is an obvious difference in information disclosure, we can also verify this

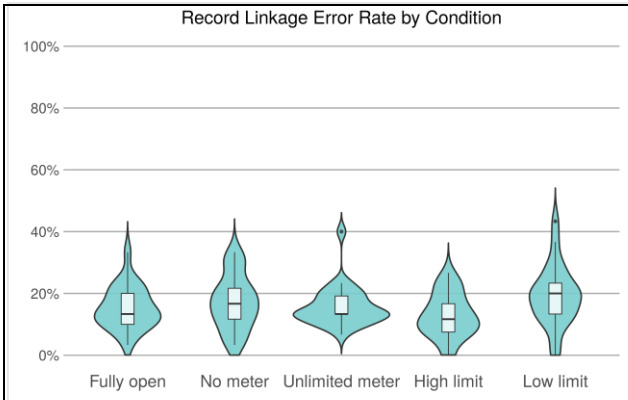


Figure 6: Percent of incorrectly linked pairs from the five conditions. Lower values indicate better performance.

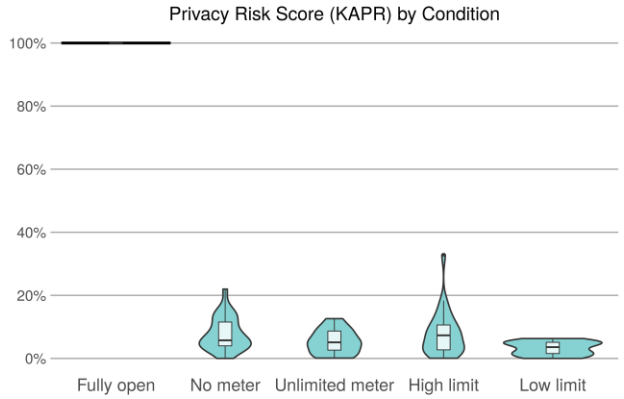


Figure 7: KAPR privacy scores for the five conditions. Lower scores indicate lower risk. Note the *fully open* condition has 100% privacy risk score due to all characters being visible by default.

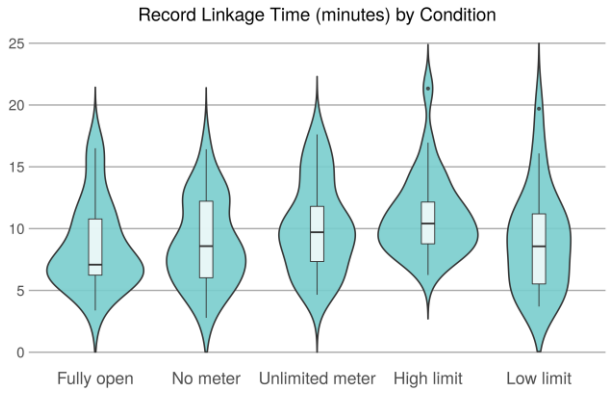


Figure 8: Time taken to complete the linkage task for the five conditions.

through inferential testing. A Wilcoxon rank sum test found a significant difference with $Z = 44$ and $p < 0.001$.

Even with such low levels of disclosure in *no meter*, the error rate did not increase significantly when compared to *fully open* (see Figure 6). A student t-test did not find a significant difference between the error rates at $t(43) = 0.77$. Though no difference was found, we cannot definitively claim that the on-demand disclosure method did not induce an increase in the error rate.

We also tested for differences in completion time between the different modes. A Student's t-test on the log-transformed completion times found no significant difference at $t(43) = 0.85$. These results support H1, demonstrating that the on-demand design significantly reduced disclosure (to only KAPR = 7.85%) with little impact on decision quality or speed.

#### 5.3. H2: Effects of feedback on the quantified privacy risk

Hypothesis 2 is concerned with the differences in information disclosure when the feedback meter is added to the on-demand clickable interface. Therefore, to address H2, we compared *no meter* and *unlimited meter*. We tested the effects of adding a feedback meter on privacy by performing a Student's t-test on transformed KAPR scores. The tests failed to detect a significant effect with $t(47) = -1.83$ and $p = 0.07$. We note that the risk score was lower when the meter was added (see Figure 7), and more data might lead to statistical differences, but further experimentation would be needed. In addition, it is worth noting that the *no meter* condition already had very low levels of disclosure leaving little room for improvement. Regardless, a Wilcoxon test on the error rates found no significant difference with $Z = 327$. A Student's t-test on the completion times also did not find a significant difference at $t(47) = -1.23$.

Thus, the study results were unable to provide evidence for H2. Both quality of decision and completion time were similar, and although adding the feedback meter to the interactive on-demand disclosure reduced the KAPR score from 7.85% to 5.33%, this difference was not statistically significant. However, the relatively low p-value ($p = 0.07$) suggests the results may be inconclusive and motivates further study, especially considering other findings indicating that people may change privacy behavior with appropriate feedback which is consistent with literature [25, 26].

#### 5.4. H3: Effects of limiting the privacy budget

Hypothesis 3 is concerned with differences in information disclosure and linkage decisions given different limits in privacy budgets. First, we compare *unlimited meter* and *high limit* to test *H3.1*, then we compare *high limit* and *low limit* to test *H3.2*. We do not compare all three conditions together because *low limit* would affect the quality score, confounding the relationship between limit and disclosure.

Hypothesis *H3.1* did not hold in the comparison between *unlimited meter* and *high limit*. A Student's t-test on transformed KAPR scores failed to detect significant differences with $t(50) = 1.46$. And a Wilcoxon test failed to detect any difference in error rate at $Z = 354.5$. Finally, a Student's t-test on log-transformed completion times found no evidence of differences at $t(50) = -1.25$.

Although disclosure levels were higher in the *high limit* condition (7.87% vs 5.33%) compared to not having a limit, it did not near the specified budget ($M = 36.7\%$, $SD = 0.81$). On average, participants used only 21.4% ($SD = 19.1$) of the given budget. Thus, the results do not provide evidence in support of hypothesis *H3*.

Providing a high limit did nudge participants to disclose slightly more in the *high limit* condition. But the study found that participants were still careful when disclosing the data. We believe this is the result of the short tutorial which emphasized opening only what was needed and participants being privacy-conscious.

However, hypothesis *H3.2* did hold in the comparison between *high limit* and *low limit*. We performed a Welch t-test on the KAPR scores, which showed evidence of differences in the risk scores with $t(35.8) = -3.46$ and $p<0.001$. For participants given the *low limit* condition, KAPR was less than half ($M = 3.22\%$, $SD = 2.12$) compared to those given the *high limit* condition ($M = 7.87\%$, $SD = 7.09$). Although this accounted for much more of the given limit ($M = 57.6\%$, $SD=36.4$) compared to the high-limit condition ($M=21.4\%$, $SD=19.1$), it was still much less than the given budget.

A Student t-test on the error rate scores found a significant difference between the modes at $t(47) = 2.62$ and $p = 0.012$. A Welch t-test on the log transformed completion times found that there was also no significant difference in completion times between the modes ($t(36.21) = 2.3$ and $p = 0.027$) at the Bonferonni-adjusted $\alpha = 0.0125$. The error results indicate that the quality of human decisions will suffer if low disclosure limits are enforced.

In sum, the interactive on-demand interface was effective in reducing disclosure to very low levels while still supporting good decisions. In addition, there is some evidence that feedback using the risk quantification may further discourage unnecessary access to PII. Limiting access via a prespecified budget may influence disclosure decisions, but more research is needed to design optimal systems to induce best behavior. Finally, the results provide further evidence that when there is not sufficient access to data, human decisions suffer.

## 6. Expert Review

We also conducted an expert review with six experts who regularly conduct record linkage and work with PII (5-10

years of experience). Experts were volunteers recruited from a professional network of people conducting record linkage studies. All experts completed an abbreviated version of the *high limit* condition used for the controlled experiment. The experts then answered questions about the potential utility and limitations of the approach and system.

In their own work, five of the experts normally conducted record linkage with full access to PII. They perceived that this system offered more privacy protection, with little to no impact on accuracy in the linkage, but may take more time. One expert had prior experience using encryption-based methods of data hiding for private record linkage with no access to PII. This participant perceived our system to have less protection and require more time compared to the encryption-based method, but to also allow for much better accuracy. He stated "I never know how well the hashing worked, or how accurate it is. It would be helpful to use this method to spot check a random sample (e.g., 5%)". This seems to agree with our goal of providing a level of access between the all or nothing that provides better accuracy than no access, but more protection than full access.

Five experts felt that the on-demand method did impact their decision making, while one did not because "I felt like I didn't need to click on most of them because my comfort level wouldn't increase". He did not think seeing more information would alleviate the uncertainty in the decision anyway. This points to the fundamental difficulty of uncertainty working with real data and affirms that the meta-data presented via the visual masks had sufficient information to support good decisions. One noted, "It works well, but it is time consuming to make the decision on whether to open the information you need".

When asked about potential benefits for this method, four mentioned privacy protection, one mentioned better accuracy of linkages, and one mentioned less fatigue of the data worker. More specifically, one expert mentioned the increased protection from the ability to accurately measure how much data was accessed (transparency) during linkage while another expert mentioned that the ability for the data custodians to limit the amount of access (budget) as being a privacy benefit. The respondent who discussed less fatigue also stated that, "Once I got used to the coding, allowing partial disclosure helped in decision making", pointing to our goal of actually improving linkage (i.e., more consistent linkage decisions) by providing better processed information for decision making in place of raw data.

When asked about specific contexts in which this system is especially useful, four stated it is useful for linking sensitive high-risk data such as health data, where privacy protection was important (e.g., "especially when linking to patient-provided data and where unique identifiers are not available"). Overall, the feedback was promising for the future potential of this direction of work, though the comments about the cognitive load for thinking about what to open suggest the need for future research, good training, and more practice.

## 7. Discussion

Research has demonstrated that information privacy is a budget-constrained problem that requires reasoning about the *tradeoff* between privacy and utility for a given context [39-41, 49]. Consequently, there is no "one-size-fits-all" solution, and there is no way to benefit from using data without taking some privacy risks. Our research tackles this difficult problem of finding the "sweet spot" between accessing PII for legitimate use while providing the maximum privacy protection as possible through the privacy by design approach.

We designed a system that reduces privacy risk through on-demand incremental information disclosure, which facilitates data work while making partial details available "as needed". This on-demand disclosure facilitates a practical implementation of the legal "minimum necessary disclosure" and accountable access requirements that are core principles of the new GDPR and HIPAA regulations making it possible to find a realistic middle ground between access to *all or no* access to PII.

From the experiment of different types of feedback and access restrictions for on-demand disclosure, the results show that all three variations were effective in increasing protection by reducing unnecessary disclosure. First, on-demand interactive disclosure (*no meter* condition) was able to significantly reduce disclosure to only KAPR = 7.85% while still being able to maintain similar quality scores. This is significantly less than what was possible with only a static display (36.7% vs. 7.85%) [20]. The prevention of unnecessary PII from being disclosed during record linkage can prevent most of the incidental identifications by people they know (e.g., neighbors, friends, co-workers) that patients are concerned about. This is exactly the local privacy that was of the most concern to patients in a survey conducted at the Mayo Clinic. For many patients, "their greatest concern about privacy actually had to do with their privacy locally ...[A] neighbor ... may still sometime be able to see [my] protected health information in the course of their work" [42, 43]. Thus, the main threat model for this work is an insider threat. It is to protect patient confidentiality by preventing someone from accidentally learning about the health status of people they know when handling PII.

Second, given the near-significant results of $p = 0.07$ for KAPR scores with the feedback meter present, this motivates interest in further study of whether quantifying the risk ahead of disclosure to inform decisions to disclose certain PII may be effective. One potential reason that differences

were not large may be due to the fact that the on demand disclosure without the meter already had very low levels of disclosure at only 7.85%. Thus, there was not much room to go lower without impacting the decision quality. Regardless of the effect of the meter on reducing disclosure, it is important to remember that quantifying the actual disclosure and sharing it with the users has a more important role. As with surveillance cameras, recording, quantifying, and displaying the risk to users has the potential to keep insiders on good behavior. One limitation of our user study is that we needed to focus on the interface and were not able to study how effective the meter was on keeping people on good behavior because the scenario we used kept everyone on good behavior. Future studies are needed to understand how much logging of computer systems, audits, and reminders of these logs might discourage bad behavior.

In addition, by quantifying and recording exactly how much risk was involved in a particular study via the meter, we can now have transparency, accountability, and communications in the record linkage process. For example, if one linkage project was able to achieve good linkage at one level, but another required much higher levels of disclosure, compliance may investigate the reason. Furthermore, with agreed-upon quantification of risks, we can now have clear conversations about what level of disclosure is appropriate at a much granular level, as apposed to limiting the options to either "access to all PII" or "no access". This conversation may include iteratively increasing or decreasing disclosure as we learn along the way.

Finally, the impact of enforcing a pre-specified limit on the disclosure was more complex. Our study clearly supported the findings from a previous study [20] that when there is not sufficient information disclosure, the quality of the linkage decision suffers (*H3.2*). On the other hand, when a sufficiently large limit was provided, participants seemed to disclose a bit more compared to the condition with unlimited budget (7.87% vs. 5.33%), though most spent only a fraction of the budget provided (21.4%). The amount disclosed was not statistically different from the *unlimited meter* condition, though the study cannot support claims for equivalence. The quality score was also similar to the *unlimited meter* condition, which may indicate that the high limit budget may be near the minimum level of disclosure needed to achieve this level of accuracy scores in the given data. This might indicate that erring on setting higher limits might be more effective since participants may still choose not to disclose the most possible, especially when they know the disclosure is transparently recorded.

The main feedback from the experts was that the system facilitated safe linkage without compromising on the quality of the results proving a good balance between the all or nothing access to PII. Some experts had concerns about the potential increase in time required for using the system. However, although there were slight increases in completion time for some interventions of our study, no statistical difference in completion time was found among the different modes. This is likely due to the fact that when we prevent users from looking at details that are not needed to increase privacy, there is a potential bonus benefit of streamlining the interface so that the users are not inundated with too much information. This is likely to reduce the time needed to complete the data task. Thus, the selective disclosure not only has the benefit of significantly reducing privacy risk, it may also have the benefit of better focused attention.

Interactive incremental disclosure that can support just-in-time decisions can be a powerful design mechanism to enhance privacy. We posit that it has the potential to have as wide an impact on privacy-enhanced systems as encryption, but inevitably the design has to be context dependent on the data task. More research is needed to understand exactly what data is needed for human decisions, when access decisions are best determined, and how to best partition access for different types of data tasks. Our findings clearly support the literature on designing better systems such as these to nudge better privacy behavior; designing systems from the beginning with privacy in mind and incorporating various interventions (e.g., education, feedback, incentives) into the system is the only way to enable safe use of sensitive data.

## 8. Conclusion

Research has demonstrated the detrimental effect of not allowing sufficient human access for data tasks [8, 11-13, 44, 45]. Errors that are not properly managed in machine-only data integration systems propagate to subsequent data analyses, which can lead to potential problems with invalid results and poor decision making. Thus, in order to obtain high quality data and bias-free record linkage, human involvement is essential to fine tune the results from automated systems (e.g., parameter settings, setting cutoff thresholds, iterative data standardization, building training datasets, validating results) [6]. Human interaction means that some data, under some suitable conditions, must be revealed to trusted persons to produce accurate linkages.

Our research provides evidence that incremental disclosure can be highly effective for ensuring legal compliance with the "minimum necessary" and accountable access requirements. Further interdisciplinary research is needed to learn the best ways to integrate these different technologies into an optimal system for privacy and utility of personal data.

## Acknowledgements

This work was funded in part by Patient Centered Outcomes Research Institute (PCORI) contract ME-1602-34486 and in part by the DARPA XAI program under N66001-17-2-4031.

## References


[1] Hye-Chung Kum, Ashok Krishnamurthy, Ashwin Machanavajjhala, and Stanley C Ahalt. Social genome: Putting big data to work for population informatics. *Computer*, 47(1):56–63, 2014.

[2] Robin E Clark, Mihail Samnaliev, Jeffrey D Baxter, and Gary Y Leung. The evidence doesn't justify steps by state medicaid programs to restrict opioid addiction treatment with buprenorphine. *Health Affairs*, 30(8):1425–1433, 2011.

[3] William H Fisher, Robin Clark, Jeffrey Baxter, Bruce Barton, Elizabeth O'Connell, and Gideon Aweh. Cooccurring risk factors for arrest among persons with opioid abuse and dependence: implications for developing interventions to limit criminal justice involvement. *Journal of substance abuse treatment*, 47(3):197–201, 2014.

[4] C Joy Stewart, Hye-Chung Kum, Richard P Barth, and Dean F Duncan. Former foster youth: Employment outcomes up to age 30. *Children and Youth Services Review*, 36:220–229, 2014.

[5] Hye-Chung Kum, Ashok Krishnamurthy, Ashwin Machanavajjhala, Michael K Reiter, and Stanley Ahalt. Privacy preserving interactive record linkage (PPIRL). *Journal of the American Medical Informatics Association*, 21(2):212–220, 2014.

[6] Hye-Chung Kum, Stanley Ahalt, and Darshana Pathak. Privacy-preserving data integration using decoupled data. In *Security and Privacy in Social Networks*, pp. 225-253. Springer, New York, NY, 2013.

[7] Hyunmo Kang, Lise Getoor, Ben Shneiderman, Mustafa Bilgic, and Louis Licamele. Interactive entity resolution in relational data: A visual analytic tool and its evaluation. *IEEE transactions on visualization and computer graphics*, 14(5):999–1014, 2008.

[8] Janet M Bronstein, Charles T Lomatsch, David Fletcher, Terri Wooten, Tsai Mei Lin, Richard Nugent, and Curtis L Lowery. Issues and biases in matching medicaid pregnancy episodes to vital records data: the arkansas experience. *Maternal and child health journal*, 13(2):250–259, 2009.

[9] Cathy J Bradley, Charles W Given, Zhehui Luo, Caralee Roberts, Glenn Copeland, and Beth A Virnig. Medicaid, medicare, and the michigan tumor registry: a linkage strategy. *Medical Decision Making*, 27(4):352–363, 2007.

[10] Francis P Boscoe, Deborah Schrag, Kun Chen, Patrick J Roohan, and Maria J Schymura. Building capacity to assess cancer care in the medicaid population in New York State. *Health services research*, 46(3):805–820, 2011.

[11] Ileana Baldi, Antonio Ponti, Roberto Zanetti, Giovannino Ciccone, Franco Merletti, and Dario Gregori. The impact of record-linkage bias in the cox model. *Journal of evaluation in clinical practice*, 16(1):92–96, 2010.

[12] Stacie B Dusetzina, Seth Tyree, Anne-Marie Meyer, Adrian Meyer, Laura Green, and William R Carpenter. Linking data for health services research: a framework and instructional guide. 2014.

[13] Partha Lahiri and Michael D Larsen. Regression analysis with linked data. *Journal of the American statistical association*, 100(469):222–230, 2005.

[14] Rob Hall and Stephen E Fienberg. Privacy-preserving record linkage. In *Privacy in statistical databases*, volume 6344, pages 269–283. Springer, 2010.

[15] Dinusha Vatsalan, Peter Christen, and Vassilios S Verykios. A taxonomy of privacy-preserving record linkage techniques. *Information Systems*, 38(6):946–969, 2013.

[16] Arvind Narayanan and Vitaly Shmatikov. *Myths and fallacies of personally identifiable information. Communications of the ACM*, 53(6):24–26, 2010.

[17] Giovanni Iachello, Ian Smith, Sunny Consolvo, Mike Chen, and Gregory D Abowd. Developing privacy guidelines for social location disclosure applications and services. In *Proceedings of the 2005 symposium on Usable privacy and security*, pages 65–76. ACM, 2005.

[18] Aritra Dasgupta and Robert Kosara. Adaptive privacypreserving visualization using parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2241–2248, 2011.

[19] Aritra Dasgupta, Min Chen, and Robert Kosara. Measuring privacy and utility in privacy-preserving visualization. In *Computer Graphics Forum*, volume 32, pages 35–47. Wiley Online Library, 2013.

[20] Eric D Ragan, Hye-Chung Kum, Gurudev Ilangovan, and Han Wang. Balancing privacy and information disclosure in interactive record linkage with visual masking. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 326. ACM, 2018.

[21] Rakibul Hasan, Eman Hassan, Yifang Li, Kelly Caine, David J Crandall, Roberto Hoyle, and Apu Kapadia. Viewer experience of obscuring scene elements in photos to enhance privacy. In *Proceedings of the 2018*

*CHI Conference on Human Factors in Computing Systems*, page 47. ACM, 2018.

[22] Serdar Çiftçi, Pavel Korshunov, Ahmet Oguz Akyuz, and Touradj Ebrahimi. Using false colors to protect visual privacy of sensitive content. In *Human Vision and Electronic Imaging Xx*, volume 9394, page 93941L. Spie-Int Soc Optical Engineering, 2015.

[23] Daphne Chang, Erin L Krupka, Eytan Adar, and Alessandro Acquisti. Engineering information disclosure: Norm shaping designs. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 587–597. ACM, 2016.

[24] Leslie K John, Alessandro Acquisti, and George Loewenstein. Strangers on a plane: Context-dependent willingness to divulge sensitive information. *Journal of consumer research*, 37(5):858–873, 2010.

[25] Alessandro Acquisti, Idris Adjerid, Rebecca Balebako, Laura Brandimarte, Lorrie Faith Cranor, Saranga Komanduri, Pedro Giovanni Leon, Norman Sadeh, Florian Schaub, Manya Sleeper, et al. Nudges for privacy and security: understanding and assisting users' choices online. *ACM Computing Surveys (CSUR)*, 50(3):44, 2017.

[26] Blase Ur, Patrick Gage Kelley, Saranga Komanduri, Joel Lee, Michael Maass, Michelle L Mazurek, Timothy Passaro, Richard Shay, Timothy Vidas, Lujo Bauer, et al. How does your password measure up? The effect of strength meters on password creation. In *USENIX Security Symposium*, pages 65–80, 2012.

[27] Ios Kotsogiannis, Ashwin Machanavajjhala, Michael Hay, and Gerome Miklau: Pythia: Data Dependent Differentially Private Algorithm Selection. In *Proceedings of International Conference on Management of Data*, pp. 1323-1337. ACM, 2017.

[28] Pierangela Samarati. Protecting respondents identities in microdata release. *IEEE transactions on Knowledge and Data Engineering* 13, no. 6 (2001): 1010-1027.

[29] Latanya Sweeney. k-anonymity: A model for protecting privacy. International Journal of Uncertainty, *Fuzziness and Knowledge-Based Systems* 10, no. 05 (2002): 557-570.

[30] Josep Domingo-Ferrer, and Torra Vicenç. A critique of k-anonymity and some of its enhancements. In 2008 *Third International Conference on Availability, Reliability and Security*, pp. 990-993. IEEE, 2008.

[31] Pierangela Samarati and Latanya Sweeney. Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. technical report. *SRI International*, 1998.

[32] Ashwin Machanavajjhala, Johannes Gehrke, Daniel Kifer, and Muthuramakrishnan Venkitasubramaniam. l-diversity: Privacy beyond k-anonymity. *22nd International Conference on Data Engineering (ICDE'06)*, pp. 24-24. IEEE, 2006

[33] Ninghui Li, Tiancheng Li, and Suresh Venkatasubramanian. t-closeness: Privacy beyond k-anonymity and l-diversity. *23rd International Conference on Data Engineering*, pp. 106-115. IEEE, 2007.

[34] Ninghui Li, Wahbeh H. Qardaji, and Dong Su. Provably private data anonymization: Or, k-anonymity meets differential privacy. *CoRR*, abs/1101.2604 49:55, 2011.

[35] Cynthia Dwork. Differential privacy. *Encyclopedia of Cryptography and Security*, pages 338–340, 2011.

[36] Pierangela Samarati. Protecting respondents identities in microdata release. *IEEE transactions on Knowledge and Data Engineering*, 13(6):1010–1027, 2001.

[37] Daniel J Weitzner, Harold Abelson, Tim Berners-Lee, Joan Feigenbaum, James Hendler, and Gerald Jay Sussman. Information accountability. *Communications of the ACM*, 51(6):82–87, 2008.

[38] Qinbo Li, Adam D'Souza, Cason Schmit, and Hye-Chung Kum. Increasing Transparent and Accountable Use of Data by Quantifying the Actual Privacy Risk in Interactive Record Linkage. Poster presentation at *Proceedings of the AMIA Symposium 2019*, Full technical report available on [arXiv:1906.03345 cs.DB] http://arxiv.org/abs/1906.03345

[39] Irit Dinur and Kobbi Nissim. Revealing information while preserving privacy. In *Proceedings of the twentysecond ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 202–210. ACM, 2003.

[40] Daniel Kifer and Ashwin Machanavajjhala. No free lunch in data privacy. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data*, pages 193–204. ACM, 2011.

[41] Tiancheng Li and Ninghui Li. On the tradeoff between privacy and utility in data publishing. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 517–526. ACM, 2009.

[42] Federal Trade Commission et al. *Innovations in health care delivery*, 2008.

[43] Daniel J Gilman and James C Cooper. There is a time to keep silent and a time to speak, the hard part is knowing which is which: Striking the balance between

privacy protection and the flow of health care information. 2009.

[44] Julia Lane and Claudia Schur. Balancing access to health data and privacy: a review of the issues and approaches for the future. *Health services research*, 45(5p2):1456–1467, 2010.

[45] Julia Lane. Optimizing the use of micro-data: An overview of the issues. *Trans. Data Privacy*, 23(3):299–317, 2007.

[46] Hanna Köpcke, Andreas Thor, and Erhard Rahm. Evaluation of entity resolution approaches on real-world match problems. *Proceedings of the VLDB Endowment*, 3(1-2):484–493, 2010.

[47] Michael Stonebraker, Daniel Bruckner, Ihab F Ilyas, George Beskales, Mitch Cherniack, Stanley B Zdonik, Alexander Pagan, and Shan Xu. Data curation at scale: The data tamer system. In *CIDR*, 2013.

[48] Hye-Chug Kum, Eric Ragan, Gurudev Ilangovan, Mahin Ramezani. Soups data and analysis (https://github.com/pinformatics/soups2019) June 7, 2019.

[49] Hye-Chung Kum and Stanley Ahalt. Privacy-by-design: Understanding data access models for secondary data. *AMIA Summits on Translational Science Proceedings*, 2013:126, 2013.

[50] In re Estate of Broderick, 34 Kan. App. 2d 695, 703, 125 P.3d 564, 570 (2005)

[51] Jennifer Guthrie. Time is running out-the burdens and challenges of HIPAA compliance: A look at preemption analysis, the minimum necessary standard, and the notice of privacy practices. Annals Health L.. 2003;12:143.

[52] Schmidt v. U.S. Dep't of Veterans Affairs, 218 F.R.D. 619, 631 (E.D. Wis. 2003), amended on reconsideration in part, 222 F.R.D. 592 (E.D. Wis. 2004)

[53] Article 5(1c) EU General Data Protection Regulation (GDPR)

[54] Cason Schmit, Kathleen Kelly, and Jennifer Bernstein. Cross Sector Data Sharing: Necessity, Challenge, and Hope*, Journal of Law, Medicine, & Ethics,* 47 S2 (2019). In press.

[55] https://www.hhs.gov/hipaa/for-professionals/faq/213/what-conditions-may-health-care-provider-use-entire-medical-record/index.html

[56] Willem G van Panhuis, Proma Paul, Claudia Emerson, John Grefenstette, Richard Wilder, Abraham J Herbst, David Heymann and Donald S Burke. A systematic review of barriers to data sharing in public health. BMC Public Health. 2014;14:1144. Published 2014 Nov 5. doi:10.1186/1471-2458-14-1144

[57] Jerry L. Hintze and Ray D. Nelson. (1998). Violin plots: a box plot-density trace synergism. *The American Statistician*, *52*(2), 181-184

# From Usability to Secure Computing and Back Again

Lucy Qin[1], Andrei Lapets[1], Frederick Jansen[1], Peter Flockhart[1], Kinan Dak Albab[1], Ira Globus-Harris[1], Shannon Roberts[2], and Mayank Varia[1]

[1]Boston University, MA, USA
[2]University of Massachusetts Amherst, MA, USA

## Abstract

Secure multi-party computation (MPC) allows multiple parties to jointly compute the output of a function while preserving the privacy of any individual party's inputs to that function. As MPC protocols transition from research prototypes to real-world applications, the usability of MPC-enabled applications is increasingly critical to their successful deployment and widespread adoption. Our Web-MPC platform, designed with a focus on usability, has been deployed for privacy-preserving data aggregation initiatives with the City of Boston and the Greater Boston Chamber of Commerce. After building and deploying an initial version of the platform, we conducted a heuristic evaluation to identify usability improvements and implemented corresponding application enhancements. However, it is difficult to gauge the effectiveness of these changes within the context of real-world deployments using traditional web analytics tools without compromising the security guarantees of the platform. This work consists of two contributions that address this challenge: (1) the Web-MPC platform has been extended with the capability to collect web analytics using existing MPC protocols, and (2) as a test of this feature and a way to inform future work, this capability has been leveraged to conduct a usability study comparing the two versions of Web-MPC. While many efforts have focused on ways to enhance the usability of privacy-preserving technologies, this study serves as a model for using a privacy-preserving data-driven approach to evaluate and enhance the usability of privacy-preserving websites and applications deployed in real-world scenarios. Data collected in this study yields insights into the relationship between usability and security; these can help inform future implementations of MPC solutions.


*USENIX Symposium on Usable Privacy and Security (SOUPS)* 2019.
August 11–13, 2019, Santa Clara, CA, USA.

## 1 Introduction

Companies, educational institutions, government agencies, and other modern organizations have been collecting and analyzing data pertaining to their internal operations for some time with great effect to evaluate performance, to improve efficiency, and to test hypotheses. While each organization's own data sets have internal value, combining data from multiple organizations and analyzing it as a single corpus is likely to provide even more value to the organizations themselves, to policymakers, or to society at large. Unfortunately, each organization's internal data sets are often proprietary and confidential, and their release may be potentially deleterious to the organization's interests. Organizations may be able to release sensitive data selectively to specific agents entrusted with its analysis. This is often costly, requires that the organizations strongly trust the agent, and presents a security risk if the data sets are improperly handled.

A cryptographic primitive called *secure multi-party computation* (MPC) resolves this tension: aggregate data may be computed and released while preserving the confidentiality of each organization's internal data. This has significant potential social benefits: MPC enables groups of organizations to leverage collective data aggregation and analysis techniques in contexts where data sharing is constrained or prevented by legal and corporate policy restrictions.

As the focus of MPC research moves from the underlying theory to the issues that affect real-world use cases, friendly user interfaces and effective communication to non-expert users is crucial for effective deployment. First, interactions with users must build trust and create buy-in to the idea of secure multi-party computation. Second, an effective user interface is especially critical for MPC applications, as it is computationally expensive (or impossible) to verify the correctness of a user's inputs after they have been encrypted. Input validation poses a problem for computations involving many parties; one party's incorrect input, whether submitted maliciously or by mistake, could skew the results of the entire computation. Multiple studies have shown that poorly

designed user interfaces hinder the success of security measures [45, 51]. A clearly designed and effective user interface can both maximize the rate of participation and minimize the chance of human errors leading to incorrect and misleading results.

## 1.1 Our contributions

This paper examines the connections between usability and cryptographically secure computing via MPC. We showcase the importance of usability within secure computing, as well as the value of secure computing in applications that calculate usability metrics in a privacy-preserving manner. We detail our experience designing and implementing a usable web-based framework for secure computing, which we call Web-MPC. We successfully deployed Web-MPC in two scenarios:

- Evaluating pay equity using the sensitive salary data of 166,705 employees from the city of Boston, MA (about 16% of the region's workforce) [14, 15].
- Measuring the rate at which member organizations within the Greater Boston Chamber of Commerce subcontract work to minority-owned businesses [42].

**Usability challenges of secure computing** The design of our framework is influenced by the interconnected usability, security, and legal requirements of our two applications, which we believe may be generalized to other scenarios. The primary benefit of our framework over prior deployments of MPC is an emphasis on the application's *usability* to drive participation within the target user community. CIOs, CTOs, human resources personnel, and lawyers from key participating organizations (along with social scientists and members of the city council that commissioned the studies) must all be able to learn and comprehend both the application itself and its underlying cryptographic properties.

We describe general usability challenges that MPC applications face in Section 3, and focus on usability challenges and solutions specific to our application in Section 4. We evaluate the effectiveness of our approach in Sections 5 and 8.

**Employing secure computing to improve usability** In the second half of this work, we describe our implementation of privacy-preserving web analytics on top of our existing Web-MPC framework. This enables the analysis of user behavior within the application without compromising any privacy goals or guarantees, and demonstrates that it is possible to measure and improve the usability of secure web applications with no privacy cost.

We discuss the importance of privacy-preserving usability data collection in Section 6, describe our implementation in Section 7, and analyze the collected results in Section 8.

## 1.2 Related Work

Usability research within the broader field of security is widespread. Countless "Why Johnny Can't Encrypt" papers have established that unless encryption is so seamless that the user cannot tell that they are encrypting, they will not bother to do so [45, 51]. Usability studies and work to incentivize use of secure platforms are common [21, 30, 46, 52]. However, these studies do not specifically tackle the difficulties in adopting MPC for a broader audience, nor do they use MPC to collect usability data.

**Usability of MPC software** There exist dozens of MPC software frameworks [27] and several successful deployments of MPC over the past decade [11, 13, 22]. These frameworks span a wide range of design choices that have usability implications for both developers and data contributors:

- Proprietary [12, 33] vs. open-source [2, 10, 17, 24, 28, 36].
- Access to low-level cryptographic primitives [8, 9, 23, 26] vs. use of programming language abstractions like data structures and formal type systems [7, 38, 44].
- Function specification in domain-specific languages [12, 44] vs. existing general-purpose programming languages [7, 47, 50].
- Whether data contributors run web servers for communication or leverage a web-based service for improved accessibility [29, 48].

Available frameworks also vary in software maturity, security guarantees, and programming APIs. Despite the widespread use of MPC, to our knowledge there has not yet been a public usability study of any application that employs MPC.

**Security and privacy of usability analytics** Within the cryptography community, previous work has been done to create privacy-preserving web analytics, with some focus on usability. However, these solutions have focused on using differential privacy [3, 18, 43], which provides fundamentally orthogonal (though compatible and complementary) privacy and security guarantees when compared to MPC.

# 2 Application Context

Our Web-MPC application was initially developed to aid a study through the Boston Women's Workforce Council. It has also been used for another initiative with the Greater Boston Chamber of Commerce. The design and implementation of Web-MPC was informed by nearly two years' worth of discussions with personnel (including CIOs, CTOs, HR executives, and lawyers from key participation organizations), social scientists, and members of the city council that commissioned the original Boston Women's Workforce Council study.

**100% Talent Compact** The Boston Women's Workforce Council (BWWC) is an initiative established in 2013 by Mayor Thomas Menino's office to measure and eliminate gender-based pay gaps [19]. In order to assess the wage gap, companies in the Greater Boston Area signed a compact promising to contribute their highly sensitive wage data across gender, race, and job categories. However, their effort was stalled due to privacy concerns over the collection of sensitive data. Employers were not willing to reveal their payrolls to a "trusted" third party and, conversely, no employer was willing to take on the role of a trusted third party due to the risk of storing or leaking such sensitive data.

As a result, we built a platform using secure multi-party computation to provide aggregate-level data statistics without collecting any company's individual data set. It has been successfully deployed in 2015, 2016, and 2017 with results and detailed analysis captured by reports through the Boston Women's Workforce Council. In the most recent 2017 deployment, the system aggregated data from 114 companies, representing 166,705 employees. This comprises over 16% of the Greater Boston Area workforce and almost $15 billion in collective annual compensation [20].

Since the application is used by HR professionals and other employees who do not have a background in cryptography, it is important for the user interface to be as intuitive as possible and to require no knowledge of the underpinning cryptographic technologies. Our application interface resembles the format of form EEO-1, which the U.S. Equal Employment Opportunity Commission requires companies to file annually. By using the familiar EEO-1 format, we aimed to improve the learnability and ease of use of our application, and to minimize errors in data submission.

**Pacesetters Initiative** The Greater Boston Chamber of Commerce (GBBC) launched the Pacesetters Initiative in January 2018 [41]. This initiative aims to enhance economic opportunities for minority-owned businesses by leveraging the purchasing capacity of medium and large businesses in the Greater Boston area [42]. A cohort of participating companies track and report metrics on their spending with Minority Business Enterprises (MBE), contrasted with their general spending across all subcontractors. The first data analysis occurred in March 2018, with the second one following a year later in February 2019. As a longitudinal study, the effort allows the GBCC to validate what effect their initiative has on equitable spending with MBEs. Given the data's sensitive nature, we partnered with the GBCC to use Web-MPC to securely and privately compute aggregate results.

## 2.1 Roles

Generalizing from our two application scenarios, we consider three types of roles in secure multi-party computation.

- An *analyst* (BWWC and GBCC in our settings) who specifies the analytics, handles some of the computational burden of calculating it, and receives its output.
- Several *contributors* who permit their private data to be used within the analytic's calculation. The number of contributors is unbounded and may be unknown in advance. In both our settings, Boston-area employers agreed to serve as contributors.
- An automated, publicly-accessible *service provider* that connects all other participants without requiring them to maintain servers or even to be online simultaneously, and that handles most of the computational burden in calculating the analytics. In our deployments, we (Boston University) configured a web server to act in this role.

## 2.2 Selection of MPC Protocol

In general, MPC assures a contributor that the analyst and service provider may only learn her data by pooling the information they receive. We rely upon passive (i.e., semi-honest) security, which informally states that if parties agree to adhere to the protocol and not collude together, then any passive attempt to glean information along the way is futile [25]. The service provider and analyst lack any clear incentive to falsify the results of the aggregation or collude to learn private input data. On the contrary, completing the data collection successfully is directly beneficial to the BWWC and GBCC (as the initiators of their respective initiatives) as well as to us as the service provider (who is incentivized to maintain a good reputation in order to deploy the application again in the future). These security protections also extend to external attackers who compromise the service provider. In more detail, MPC guarantees that read attacks against the service provider yield no private input data. Our implementation and MPC protocol also guarantees that inputs cannot be linked to the original contributor, parties that did or did not submit cannot be identified, and that the number of users does not need to be determined in advance.

We surveyed existing MPC implementations and their designs at the beginning of our effort [32]. None of the existing implementations at the time sufficed for our purposes. Some of them required the analyst to configure a public-facing web server, whereas others failed to provide the accessibility, asynchrony, auditability, resubmission, or other usability requirements listed in Section 4.2.

Instead of using existing frameworks, we opted to create a simple MPC protocol that was easy to implement without errors, straightforward to explain to users who are not domain experts, and adaptive enough to handle an *a priori* unknown number of participants. The protocol uses a variant of additive secret sharing in which random masks are added to each company's private inputs, as shown in Figure 1. The service provider (Boston University) then computes the aggregate

sum of the masked inputs while the analyst (BWWC) receives the masks. The analyst is then able to subtract the aggregate of the masks from the aggregate of the masked values to get the aggregate data. This protocol is detailed in Appendix C.

## 2.3 Application Versions

Our Web-MPC application has gone through multiple iterations over time. There are two versions we will consider in our discussions, and we will refer to them as V1 and V2. V1 refers to first iteration of the application that uses additive secret sharing. The V1 data submission page is displayed in Figure 4. V2 refers to the current iteration of the application, after changes were made to the user interface based on the heuristics evaluation detailed in Section 5.1. Instead of additive secret sharing, V2 uses Shamir's Secret Sharing to support richer analysis (as discussed in Section 4.1). The most recent version also enables the creation of smaller participant subsets, called cohorts, within a session. For the 100% Talent Compact, these cohorts will consist of companies grouped by industry; the Pacesetters Initiative divides participants into cohorts based on prior participation. Aspects of the V2 user interface are captured in Figures 6, 8, 7, and 9. The differences between the V1 and V2 iterations of the platform are discussed in greater detail in Appendix B.

## 2.4 Deployment

The protocol and software application described in this section were deployed successfully by the BWWC [6, 35] and by the GBCC. We split each deployment into two phases: (1) a dry run on innocuous data and (2) a live analysis over sensitive data during which our team remained on-call to ensure any potential technical issues or usage questions were addressed.

**Training Sessions** The application's learnability, familiarity, and ease of use are critical to minimizing user error and to achieving a successful secure deployment. To preserve privacy, we could not help participants enter data into the user interface or allow them to ask us questions dependent on their data during actual deployments. This concerned us: we felt that we only had one chance to introduce MPC to participants. If even one participant entered erroneous data and the output was unsatisfactory, they might blame the technology (and switch to something less secure or not participate at all). To reduce this risk, we conducted dry runs involving fictitious data (provided to participants) in which they could ask us questions, and become familiar with the application interface and submission flow.

A dry run of the deployment served the purpose of familiarizing participants with the protocol, process, interface and requirements. We distributed an Excel spreadsheet that exactly matched the browser data entry interface, and ensured interoperability between the two. Contributors could use this spreadsheet to prepare their data and to verify that their browser allowed them to copy and paste data directly into the web application. The entire workflow was demonstrated via a live WebEx session that all participants could join. We initiated a mock collection, shared the session key, and encouraged all contributors to submit random data. This WebEx session was recorded, uploaded to YouTube, and shared with all participants so they could review it at their own pace. The training sessions provided the contributors with opportunities to ask questions, and allowed us to discover technical issues contributors might encounter (*e.g.*, using an outdated browser) without the risk of leaking information about their inputs.

**Live Deployments** In both deployments, the analyst was able to perform MPC jointly with our server to privately compute the aggregate across all contributing parties, and to delete their private key (in effect erasing the input data).

## 2.5 Mechanical Turk Usability Study

Prior to recent deployments, we conducted a usability study using Amazon's Mechanical Turk (MTurk) Platform to evaluate the success of different iterations of our application. We utilized MTurk users, rather than data contributors involved in either the BWWC or the Pacesetters deployments, in order to obtain data from a larger audience that is not already familiar with versions of the interface. Current data contributors have either already seen previous iterations of the platform or have received training during which they had the opportunity to ask questions. By using MTurk, we could assess whether our tool is usable with limited instructions and no prior training. During the study, MTurk users used our application to submit data before completing a System Usability Scale (SUS) questionnaire [16]. MTurks users were split into three groups. The first interacted with V1; the second and third interacted with V2, but were asked to enter data manually or via a spreadsheet. We describe the SUS and its results in Section 5.2. Additionally, during this study we collected usability data securely via MPC in the background as a proof of concept.

# 3 Usability Challenges in Deploying MPC

MPC introduces unique usability challenges. Target users are not domain experts and are unfamiliar with this technology; their willingness to use an application depends on their confidence that MPC protects their sensitive data and guarantees compliance with data sharing requirements. Also, the inherent privacy-preserving properties of MPC make it difficult or impossible to identify spurious or erroneous contributions that might compromise the overall analysis. Thus, the application's learnability, familiarity, and ease of use are also critical to minimize errors.

## 3.1 Inspiring Trust in MPC

Unlike more popular cryptographic primitives (*e.g.*, end-to-end encryption), MPC is not yet ubiquitously used in practice. MPC's guarantees and terminology are not widely circulated within non-technical or semi-technical contexts. As a result, HR personel, lawyers, CEOs, and end users are less likely to be aware of MPC guarantees or to have high confidence in it. This puts an additional burden on MPC application designers to communicate the guarantees of MPC to the relevant stakeholders and inspire confidence in its security.

This cannot be achieved solely by relying on mathematical and cryptographic proofs. Such proofs are not accessible to the wider population. Furthermore, they may not be convincing to someone that does understand what a proof entails. In our experience, analogies, examples, and concrete demonstrations of MPC play a large role in this endeavor.

Some contributors still require that the various compute parties within an MPC application sign a non-disclosure agreement governing data submitted by the contributors. This can be attributed to a lack of understanding or confidence in MPC guarantees, as well as familiarity with NDAs and their use to mitigate liability. However, due to the way that MPC works, we believe participants would benefit more from the creation of a different legal construct: a "non-collusion agreement" with enforceable civil penalties.

## 3.2 Correctness and Participation Trade-offs

When building usable MPC platforms, designers must negotiate an inherent trade-off between the participation rate, the correctness of the aggregate output data, and the security of the input data. With an increase in contributors, the chance that at least one contributor provides incorrect data increases; in other words, increased participation adversely impacts correctness. Simultaneously, participants are more likely to participate if they have confidence that the computation will be correct (and, thus, useful).

**Error Sensitivity** Unlike traditional computation platforms, the nature of dealing with private inputs under MPC makes error recovery tedious (if not impossible). MPC does not allow any single party to look at input data, or to analyze it manually. This makes it difficult to detect and correct invalid input data and to remove outliers. It is also difficult to use contributor-specific context that could normally inform an analyst of potentially incorrect data. For example, a publicly traded multinational company that submits an input indicating it employs only five individuals is likely to represent a mistake, but this cannot be determined easily without seeing the individual inputs. Detection and correction logic can be encoded to run under MPC. However, this increases performance overhead and may require that all such logic be formalized and written down without prior knowledge of the input data's characteristics.

All these issues make it critical for the application to detect errors *before* contributors submit erroneous data and taint the aggregate results. This, in turn, necessitates that the application be easy to use and to learn, and that it allows contributors to review, correct and resubmit any erroneous data. Contributors cannot contact application maintainers during deployment, as this may inadvertently reveal information (for example, certain errors may only occur when the input meets certain conditions). We attempt to increase the contributors' familiarity with the application and to decrease their need for support during its deployment by holding training sessions (as described in Section 2.4).

**Benefits of Participation** We also note that output privacy (informally speaking) increases with the number of contributors, as the output is an aggregate of the inputs (e.g., in the pathological case of a single contributor providing input, that input is necessarily leaked by the output). Thus, the simplicity and accessibility of the framework, as well as the comprehensibility of the underlying cryptographic tools indirectly contribute to the overall security of the protocol by encouraging participation.

# 4 Usability Challenges within our Use Cases

Our application requirements and deployment scenario posed specific usability challenges, in addition to the general usability challenges described in Section 3. We describe our solutions to these challenges throughout this section.

## 4.1 Communicating MPC

In order to convey MPC's security guarantees to non-experts, we devised various analogies for additive secret sharing that constitute a possible workflow they can replicate on their own. While our evidence is purely anecdotal, we surmised that some explanations worked better than others. In an early example, we attempted to demonstrate the process of splitting values into shares, which was visually represented by dividing bars into smaller bars and then reassembling them with the pieces of others. This limited us to using only positive values (negative space is difficult to represent), and in turn falsely gave some participants the impression that we would leak the lower bound of their data. In another example, and partly as a response to the lower bound question, we used clocks with the summation of a random value to explain the concept of finite fields or modulo. This in turn raised questions about the process of joining multiple clocks together, and how the actual data would not get lost in the process.

One analogy that did appear to resonates with our target audiences, and could be explained outside of a presentation, is describing the process as *lying about your salary* to one party,

and letting another party know *by how much you lied* (*i.e.*, an offset); neither number reveals your actual salary to either party. But if multiple contributors give their lies and offsets to the two parties, each party can tally what they have (either lies or offsets) and then subtract one sum from the other to obtain a total of the original values.

The analogies are not meant to explain intricate details of MPC, but to give the audience confidence that it is possible to compute a function without revealing private inputs to those performing the computation. We estimate that about 500 people viewed our presentations, with attendees present at training sessions, conference talks, academic events, corporate conference calls, and other venues. While we cannot claim to know the total number of people whose minds were changed as the direct result of our presentations, both (1) personal feedback after the sessions and (2) a marked uptick in BWWC Compact Signers who indicated their willingness to contribute data indicate the effectiveness of our communication efforts.

We accompany our explanations with a diagram of our protocol that includes the public-key cryptography required to allow one enabling party to handle all communications. Figure 1 accurately reflects the MPC protocol used in the first iteration of the platform, as additive secret sharing met the basic analytic needs (*i.e.*, averages) and was straightforward to explain. Our implementation was open-source and available for anyone to audit. In later deployments, analysts specified more complex analytics (*e.g.*, deviations and longitudinal analysis of specific cohorts) that required a general-purpose MPC library that relies on Shamir's Secret Sharing [49].

Shamir's Secret Sharing is more complex than the additive secret sharing scheme described initially (*e.g.*, it relies on properties of polynomials over a finite field). We found that non-experienced personnel were willing to have more confidence in this scheme after being exposed and familiarized with simpler variants like additive secret sharing. They were more receptive to the idea that any function can be computed on private data once they understood how *some* functions could be computed. This appreciation increased the willingness of several participants to contribute sensitive data despite an initial unwillingness to do so.

## 4.2 Usability Requirements

Our application scenarios involve individuals with a wide range of technical backgrounds utilizing computing resources that are outside our control and governed by a variety of organizational constraints. Thus, application usability is critical. For a starting point grounded in the literature, we referred to the five usability components defined in seminal work by Jakob Nielsen and other human factors practitioners [39]. However, we found that these five components were insufficient to fully and faithfully characterize our usability requirements. This is due to the distinct usability challenges associated with deploying MPC both in general and in our specific application. Instead, we present a novel categorization of usability requirements that we believe is more suitable to our MPC application. Where possible, we indicate how our category relates to those defined by Nielsen. We recognize that our requirements may not be well-suited for evaluating the usability of privacy-preserving software in general; that is a broader issue outside of the scope of this paper.

**Error Minimization** Errors are particularly problematic in our setting, as MPC's encoding of data prevents us from detecting or sanitizing bad inputs. In addition to allowing users to copy and paste the data (rather than use a more error-prone manual entry process), we proactively provide feedback to warn users about missing or spurious (*e.g.*, out of range) data prior to submission (see Figures 4 and 8). We also compel users to consciously confirm that their submissions do not have errors by requiring that they click a checkbox before submission attesting to this fact. Additionally, we permit contributors to resubmit their data if they discover that a previous submission was corrupted due to human error or software failure. In general, supporting re-submission influences the design of the underlying MPC protocols, since not all MPC protocols can support it without modification.

**Asynchronous Participation** In traditional usability, this can be viewed as a component of *subjective satisfaction* as defined by Neilsen. However, it is particularly important for our MPC application. We used software support for asynchronous participation in order to satisfy the logistical challenge of scheduling a data analysis effort that involves numerous enterprises of various sizes. This requirement has significant implications when employing MPC because it dictates which MPC protocols can be used. In fact, this requirement was a factor in our decision to design a new MPC software framework. When using our software, contributors only need to be online while entering their data, and the analyst only needs to be online at two points in the protocol: to start the process and to compute the analytic. The analyst additionally needs to store and safeguard one piece of information (an RSA private key) between these two points in time. Finally, the analyst has the ability to see how many submissions have taken place on their interface; this feedback reassures them that the application is being used successfully by contributors.

**Ease of Use** This umbrella component includes the *learnability*, *efficiency of use*, and *memorability* requirements as defined by Neilsen. These requirements have significance in MPC use cases only in that addressing them can reduce submission errors; their significance otherwise is primarily determined by their importance in informing the design of *any* data entry system. Our application meets these requirements: it relies on familiar web interfaces, and requires (on the part of the contributor and of the analyst) no setup process,

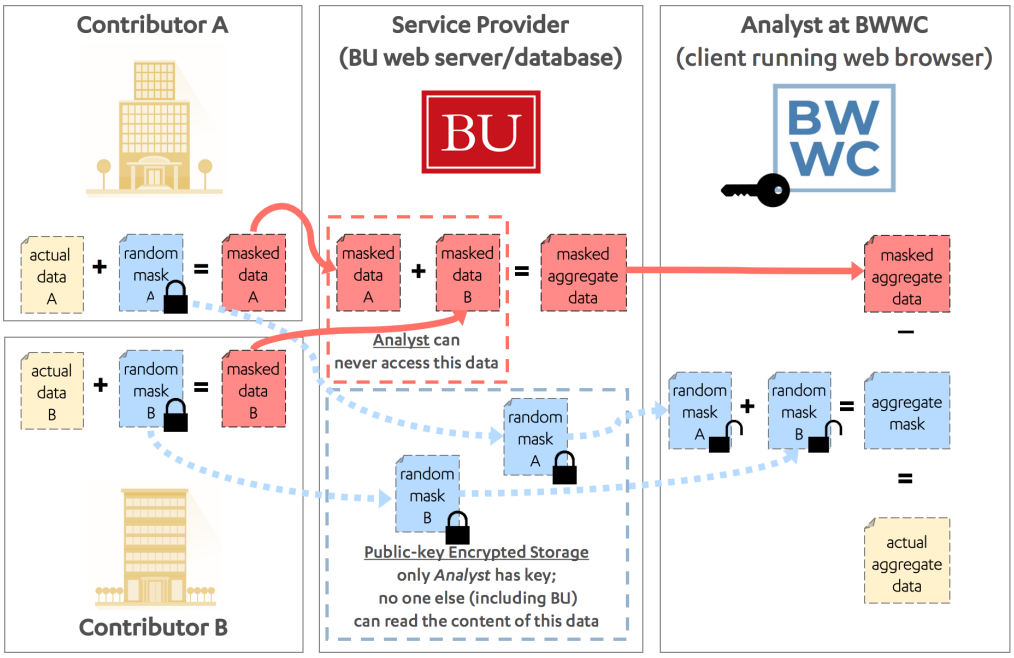



Figure 1: Slide detailing the MPC protocol. This was used in explanations to HR employees, lawyers, CEOs, and so on.

no specialized software or hardware, and no management of public internet addresses. Additionally, we provide users with an Excel spreadsheet to complete ahead of time; this sheet requires the same information as the web application, in the same format. Users can simply copy and paste the information from one source to another, or import the spreadsheet directly into our interface. Even when the UI is improved, the spreadsheet template and design remain fixed between deployments, improving the memorability of our system (especially because companies use it to recurringly provide their salary data).

**Comprehension and Trust** MPC technology can be difficult to explain to non-experts. It is critical to ensure contributors understand and trust the security guarantees of MPC, as this can increase the rate of participation. This challenge is unique to privacy preserving technologies as described in Section 3.1, and is not emphasized in Neilsen's requirements. Section 4.1 describes our approach: we initially chose a protocol that is easier to explain, accompanied by analogies and visualizations to improve comprehensibility for non-experts. Finally, our implementation is open-source to inspire trust.

## 4.3 Workflow Design

Given that the purpose of this application is to allow a group of non-domain expert participants to execute a session of a secret sharing scheme, we had to make several design choices that increase the usability of the system without compromising its security.

**Experience of Contributors** The application automates almost all portions of the protocol. The analyst must distribute the session identifier and participation codes to the participants (*e.g.*, via email). This ensures that the participants' experience is simple and error-resilient: they can simply click on a link (without entering participation codes or long identifiers manually). Because the analyst sends these links via an external channel, the service provider cannot see any correlation between participation codes and individual contributors. The service provider does not allow the analyst to see which participation codes were used to submit and which were not; only a count of contributions and their submission times are shown.

Participants must enter data either through manual entry or by pre-populating a formatted spreadsheet that has been provided to them and importing the file into their web browser. Since realistic scenarios involve not one but a collection or table of labeled numeric quantities from each participant, the software application actually implements the protocol in parallel on multiple labeled fields within a table.

**Experience of Analysts** The analyst starts a session by clicking a single button in the analyst interface and saving the 2048-bit private RSA key (unique to that session) to their hard drive. This interface also enables the creation of cohorts and the generation of participation links for contributors. Finally, it includes a session tracker that displays an anonymous participant submission history, as depicted in Figure 6. Their second interface, the final unmasking page, computes the final aggregate data upon successful submission of the analyst's

private RSA key associated with the session.

Due to the role of the analyst in our system, the usability of their interface is not as critical as the usability of the contributors' interface. This is partly due to the fact that the analyst does not provide numerical inputs, and thus is not responsible for any error-prone input tasks that can affect the quality of the computation. Additionally, analysts had constant interactions with us as we developed the application, and thus have more experience navigating and using it.

If too few participants have submitted their data (the minimum number of participants can be configured) the service provider will not allow the analyst to compute the final aggregate data. Once the final aggregate is computed, it is displayed in the same familiar table format as the input table presented to individual participants.[1]

## 5 Usability Evaluation

We conducted two kinds of common usability evaluations, a Heuristics Evaluation and a System Usability Scale (SUS). We used both to evaluate our application's flow and interface. The results of the Heuristics Evaluation informed the subsequent redesign of our V1 interface into the V2 version. We discuss these evaluations and their results throughout this section.

Figure 4 illustrates the original web interface (prior to changes made based on the heuristics study) used for the BWWC study[2] as it appears within a web browser to each contributor. The participant interface provides a familiar spreadsheet table that an end-user can fill with data either manually or by pasting the data from another application. The email address is hashed on the client-side and this hashed value is used only as an index into the server database, allowing each participant to submit more than once in a session (overwriting their previous submissions).[3]

### 5.1 Heuristics Evaluation

After deploying our web application twice for the BWWC, a heuristic evaluation was conducted on V1 (displayed in Figure 4) to serve as an iterative design tool for addressing usability concerns. Heuristic evaluations involve having a set of evaluators examine the web application to judge its compliance with recognized usability principles (known as "heuristics"). More information can be found in Appendix A.

[1] It is the responsibility of the analyst to destroy their local copy of the private key after retrieving the result if this is the agreed-upon procedure. That is: assuming secure erasures, we achieve forward secrecy when the protocol is composed.

[2] The client application can be viewed at `https://100talent.org`.

[3] After submission, data remains visible in cleartext in the participant's browser so that any errors can be identified and a fixed set of inputs can be resubmitted. We relied upon briefings to inform participants about the post-submission encryption process.

**Results** A total of 32 issues were identified. Each issue, along with its categorization and average severity rating, are shown in Tables 4 and 5. Two issues ("There is no indication as to whether the email address is valid" and "There is no email confirmation indicating that data was submitted") directly highlight the trade-offs between usability, correctness, and security discussed in Section 4.2. The security requirements preclude any server-side validation of the identity of a data contributor (because we impose the requirement that no individual participant can be identified by any part of the application) or any communications from the server to the client via another channel that requires their identity (such as email messages). At the same time, an error rate of zero is required, since any data entry errors lead to incorrect aggregate results. Thus, it is still necessary to allow users to resubmit data if they notice they made a mistake. The compromise was to create a unique identifier corresponding to the clients' submission.

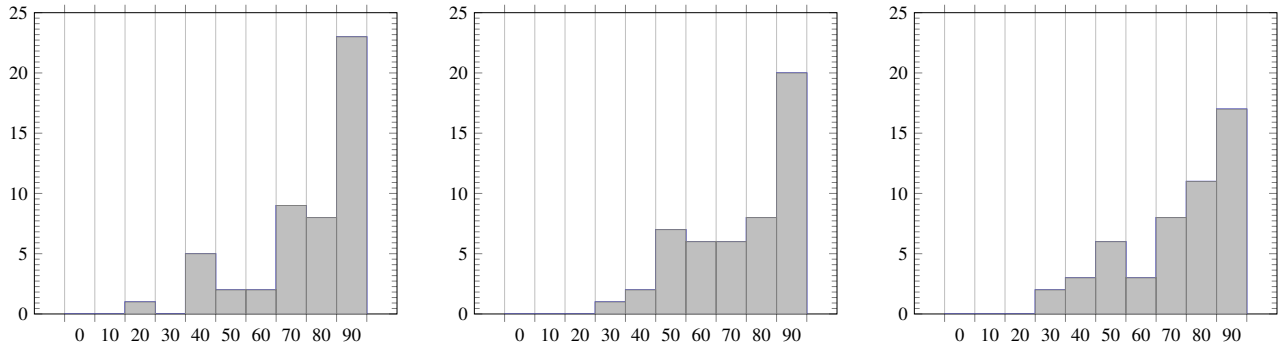
Ten issues (2.1, 2.2, 2.3, 2.5, 4.1, 4.2, 4.3, 4.4, 4.5, and 6.2 in Table 4) are difficult to correct because the layout and format of the web application is based on the Equal Employment Opportunity Commission's (EEOC) Employer Information Report EEO-1 [1], which large-employer organizations are required to complete and file annually. Changing the web application to address these ten issues would lead to a mismatch with the EEOC Report, potentially confusing users and reducing memorability.

Some issues (1.5, 4.2, 4.3, 4.4, 4.5, 6.1, and 6.2) were addressed during in-person training and/or within training materials . Other issues were deemed out of scope for this project (3.1, 7.1), or no longer relevant with the introduction of new workflows.

**Interface Redesign** We used feedback from the heuristics evaluation to redesign our application's data submission interface prior to the 2017 deployment. The redesign divided the application into four distinct steps, each visually and logically separated by a card layout that clearly delineates varying steps in the submission process, and addressed the remaining issues from the evaluation. More details can be found in Appendix B.

### 5.2 System Usability Scale

Prior to our recent deployments, we conducted a usability study on Amazon's Mechanical Turk (MTurk) Platform. The users were all directed to the Pacesetters Initiative data collection platform and asked to fill in the nine data fields with numerical data. Afterwards, users completed the System Usability Scale (SUS) questionnaire [16], comprised of 10 questions evaluating the application. Using the SUS allows us to measure our application's perceived usability with a relatively small sample size. It is also advantageous because it is quick for the participants to complete.




(a) Group 1 SUS Scores (b) Group 2 SUS Results (c) Group 3 SUS Results

Figure 2: Histograms of SUS results, binned with width 10 on the x-axes and number of respondents on the y-axes.

**Study Methodology** To successfully submit data on the web page, users were required to:

1. Load the page
2. Fill out the nine fields of the form with numerical values of their choosing (either manually or by uploading a spreadsheet, depending on the experimental group)
3. Check a box to indicate the submission is correct
4. Correct any errors indicated by the user interface after the verification box is checked
5. Press the button labeled "Submit"

Unlike participants for the BWWC and GBCC initiatives, Mechanical Turk participants did not receive any prior training on the platform or an explanation of MPC.

**Experimental groups** Data sets were collected from three experimental groups:

- ***Group 1***: Users who interacted with V1 of the web page; V1 has the original appearance and interface of the platform as first deployed in 2015, configured with the data fields from the Pacesetters initiative as displayed in Figure 5. The users had to manually enter data into each of the cells as there was no import option.
- ***Group 2***: Users who interacted with the current design, V2, of the web page. These users were instructed to manually enter data into each of the cells. The drag-and-drop import box, as depicted in Figure 7, was removed to avoid confusion.
- ***Group 3***: Users who interacted with the current design, V2, of the web page, and were asked to fill out a spreadsheet on their local machine and upload it to the page via either the drag-and-drop or the file import feature.

For each group, 50 participants were recruited on the Amazon Turk Platform. Participants were paid $2.50 for completing the task, which at the beginning of the study was estimated to take 10 minutes. Previous studies have noted demographic differences between MTurk workers and the general population [34], specifically that MTurk workers tend to be younger and have lower incomes than the general population. Users are likely more familiar with common website interfaces, which may skew the SUS scores collected slightly higher than they might be for users of the BWWC and GBCC applications.

**Results** The results of the SUS Survey are intended to be a general indicator of a system's usability, rather than a diagnostic of specific system successes or failures. Histograms of the SUS scores by group are displayed in Figure 2. The mean and median SUS scores are displayed in Table 1. Group 1 scored highest, and all scores were above average compared to an analysis by Bangor et al. in 2009, which reports a mean SUS score of 68.2 for the over 3400 web pages surveyed [5]. However, one might speculate this comparison to Bangor et al. is limited due to improvements in web technology and design standards, as well as increased familiarity with web interfaces among the general population.

| | Group 1 | Group 2 | Group 3 |
|---|---|---|---|
| Mean | 80.4 | 78.3 | 76.9 |
| Median | 85.00 | 83.75 | 82.50 |

Table 1: The mean and median of SUS scores for each experimental group.

In order to determine if there was a significant difference between the SUS scores per group, a Kruskal-Wallis one-way analysis of variance was performed on the SUS scores in each group [31]. The resulting $p$-value of 0.474 is not sufficient to reject the null hypothesis that there is no difference. Thus, although more recent user interfaces that we tested (Groups 2 and 3) scored slightly lower on the SUS scale, the scores do not indicate a statistically significant decrease in overall usability.

## 6 MPC for Improving Usability

Collecting user data on websites is an accepted best practice for understanding and improving their design and usability. Web analytics platforms such as Google Analytics and Mixpanel provide valuable feedback on application usage and facilitate data driven methods for improving usability features, accessibility, and design efficacy. These platforms are compatible with almost all web pages and provide valuable usage data, but also collect a multitude of identifying factors about users that expose information to both the site administrator and third-party platform providers [4]. While collecting information about how people interact with an application is helpful for interface enhancements, it is naturally at odds with the goals of an application designed to protect user privacy.

Our goal was to introduce a collection method for web usability metrics and usage data that did not compromise the privacy goals of the original application. Previous deployments of Web-MPC lacked a method to analyze how users are interacting with the application, as we believed collecting this data could undermine the privacy of the users and their trust in the system.

## 7 Collection of Usability Data via MPC

Secure web analytics data collection via MPC was integrated into the implementation of Web-MPC. Client-side analytics were collected through the existing MPC protocol, meaning that no new cryptographic protocols or libraries were required to collect and analyze the data. The usability data sets were masked in the same manner as the application data (salary data for BWWC and spending data for GBCC). The masked values are also sent to the *service provider* while the masks were encrypted and sent to the *analyst*. The metrics are revealed in aggregate form to the analyst in the same way as the application data sets that are contributed, meaning that no piece of information can be tied to a specific user. No modifications were made to the submission workflow itself or the underlying protocol detailed in Appendix C.

### 7.1 Usability Metrics

The following metrics were measured on the client side and submitted via MPC. Metrics are captured either on entry time (when a user completes a field) or on submission (when a user clicks the submit button); all metrics are only communicated to the server upon successful submission.

*Time spent* measures the number of milliseconds a user spends on various areas of the user interface. It is captured for each card of the layout: session area (cf. Fig. 7, table area (cf. Fig. 8, submission area (cf. Fig. 9). In the current UI, time spent is also tracked for the entire table area and the review area.

*Data Prefill* is a metric that measures if a user filled out the data using a spreadsheet and then submitted it by either importing or dragging and dropping the spreadsheet.

*Validation errors* refer to any error encountered by the participant while interacting with the user interface as enumerated below.

- ***Session info error*** occurs when a user enters an invalid session key or participation code.
- ***Empty cell error*** refers to any occurrence in which the user leaves a table cell empty during manual entry.
- ***Invalid cell error*** refers to any occurrence in which the user does not enter integers into the table cell during manual entry.
- ***Submission cell error*** measures whether there are any remaining empty cell or invalid cell errors when a user clicks the submit button. It does not count the number of errors remaining. This metric increments by 1 each time a user attempts to submit with remaining errors.
- ***Unchecked error*** occurs when a user attempts to submit data without first verifying that they have double-checked the values they have entered.
- ***Server error*** is captured if a user attempts to submit but is unsuccessful and encounters a status of 0 or 500, indicating a server-side error.
- ***Generic submission error*** is captured if a participant attempts to submit but is unsuccessful and receives a status of anything other than 200, 0, or 500.

## 8 Results and Analysis

The usability study ran over a three-day period and had 150 participants from Mechanical Turk. Of those 150 participants, 143 were able to successfully submit data with a total number of 200 submissions made (including re-submissions). Although 7 participants did not successfully submit data, all 150 participants completed the SUS survey (discussed in Section 5.2).

We emphasize that the results and success of the user study serve as a proof-of-concept for MPC-enabled web-analytics collection, and demonstrates the general feasibility of MPC as a tool to inform and improve application usability.

### 8.1 Usability Metrics Results

The usability metrics collected during use of the application are presented in Table 3, which contains the breakdown of errors encountered by users, and Figure 3, which presents the average time spent on the page by the different user groups.

| Group | # Participants | # Submissions | # Re-Submissions |
|---|---|---|---|
| 1 | 50 | 72 | 22 |
| 2 | 49 | 69 | 20 |
| 3 | 44 | 59 | 15 |

Table 2: Total number of successful participants and submissions by group.

| Validation Error | Group 1 | Group 2 | Group 3 |
|---|---|---|---|
| Empty cell | 0.56 (28) | 0.20 (10) | 0.07 (3) |
| Invalid cell | 0.04 (2) | 0.02 (1) | 0.00 (0) |
| Submission cell | 0.08 (4) | 0.16 (8) | 0.05 (2) |
| Unchecked | 0.02 (1) | 0.04 (2) | 0.00 (0) |
| Session info | 0.00 (0) | 0.00 (0) | 0.00 (0) |
| Server | 0.00 (0) | 0.00 (0) | 0.00 (0) |
| Generic submission | 0.00 (0) | 0.00 (0) | 0.00 (0) |

Table 3: Average errors per user for each experimental group. The number of participants was 50, 49, 44 for Groups 1, 2, 3, respectively. The absolute number of total errors encountered by group is listed in the parenthesis.

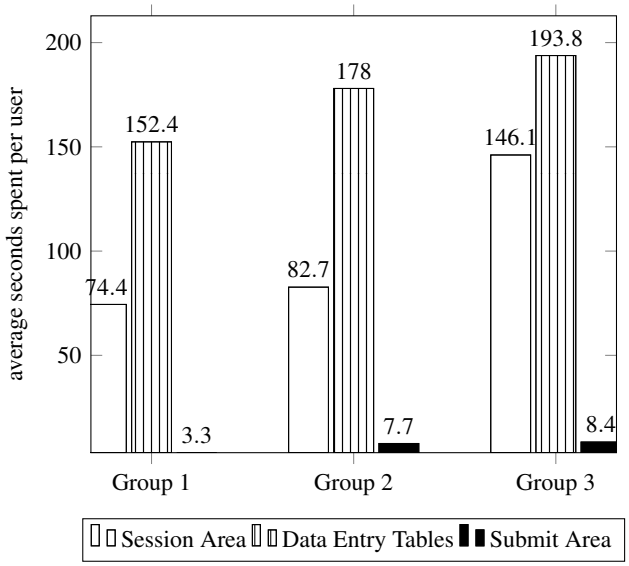



Figure 3: Average time spent per participant in seconds on the individual data entry tables by group.

## 8.2 Analysis

First, the results of the study confirm the feasibility of using MPC to understand and improve application usability. This could be achieved by introducing an MPC library to an existing web application in order to collect data in a privacy preserving way, or, more reasonably, by adding usability data analysis features to applications that already employ MPC.

With regard to Web-MPC, the results of the study suggest that the usability enhancements made to the application led to fewer data entry errors, even though some participants experienced difficulties with importing pre-filled spreadsheets. Only 21 of the 50 participants in this cohort chose to upload their pre-filled spreadsheets, which may account for the lower SUS results. It is expected that there would be fewer submission cell errors for Group 3, as their data would have been already filled out ahead of time. This should reduce the number of errors due to an empty cell or an invalid cell, as well. Even though only 21 participants imported spreadsheet files, there were only 2 total submission cell errors; this implies that fewer users encountered difficulties at submission time.

Although participants in Group 1 had a slightly higher SUS score, they encountered more errors in almost all categories compared to both Group 2 and Group 3. The version of the interface they used highlighted cells with errors in red but did not have tooltips explaining what the error was, which may have made it more difficult for participants to understand how to fix their errors. As shown in Figure 8, the V2 submission page provides additional instructions for participants compared to the V1 page. The additional instructions explaining what red and yellow cells indicate may have assisted these users in fixing their errors and could have prevented them from making the same mistake in subsequent cells.

## 8.3 Limitations

A key limitation of the process for gathering usability metrics is that information is only collected upon submission, when the secret sharing scheme is initiated. Thus, we cannot obtain metrics from participants who may have had too much trouble with the platform to successfully submit any data. In Group 1, all 50 participants successfully submitted data and metrics were captured for all of them. For Group 3, 6 participants were unable to complete the data submission process. Their contributions were therefore omitted from the final aggregate results. To account for this, the SUS survey was completed by all participants regardless of their ability to successfully contribute data.

Another limitation is that we are not able to link usability metrics back to individual users (as per the security guarantees of the MPC protocol). While this ensures user privacy, it limits the ability to gauge variance between the results. It is possible that all errors are produced by a single outlier participant in each group.

For each of the experimental groups, there were multiple resubmissions of data. Each time a re-submission occurs, the data corresponding to the participant is overwritten. This includes all of the data capturing the participant's submission errors and usage data. Thus, we only have the usability metrics from each participant's most recent submission, omitting their initial behavior with the platform. This may understate the number of errors encountered because we cannot analyze client-side usability metrics from previous data submissions.

As empty and invalid cell errors are captured upon manual entry, the results for Group 3 emerge either from participants

who manually submitted or from those who attempted to fix cells, adding errors after importing a spreadsheet.

## 9 Future Work

The integration of usability metrics into a MPC protocol allows us to inform ongoing iterations of the user interface design using data from both usability studies and real-world deployments. We plan to collect usability data from participants in future deployments of the application with the BWWC and the GBCC. We also plan to continue expanding the number of usability metrics we collect.

As previously discussed, client-side usability metrics are only secret-shared upon successful data submission. We would like to gather usability data from all submission attempts. This method for gathering usability metrics has helped us mitigate the tension between the need to gather data on user behavior and the need to ensure the privacy of user activity and inputs. We intend to generalize this solution into a standalone plugin that can be used to allow other web services to obtain web analytics in a privacy-preserving way.

## 10 Lessons Learned

In order for privacy-preserving applications to be deployed and used to analyze sensitive data, the relevant stakeholders need to be convinced of the privacy guarantees of the underlying protocols while being insulated from the nuances of their implementation. Although there are a broad array of MPC protocols available, our initial use of additive secret sharing enabled relatively straightforward explanations of MPC. Demonstrating how MPC works using a simple, concrete function increased confidence among potential users that more complicated functions may be computed securely.

The user interface was designed in a manner that was familiar for participants, which also aided their ability to participate. Although participants may not have initially been familiar with the cryptographic techniques used, they were familiar with the task of filling out cells in a spreadsheet. The rest of the details of the protocol were abstracted away.

As our experimental results demonstrate, incorporating the improvements identified via the heuristic evaluation reduced the amount of human error during data entry. This shows that deployments such as these may be well-served if they are executed as joint collaboration between security engineers (who can design a technically sound framework) and human factors experts (who can improve UI/UX features before, during, and after deployment). Such collaboration between application developers and human factors experts is particularly important for privacy-preserving applications with which users have limited opportunities to interact: the applications must be usable even when users encounter them for the first time.

Gathering user behavior data in privacy-preserving applications may leak information, or create side channels for identifying or linking inputs to clients. However, MPC itself can be used to gather these metrics while still maintaining the guarantee that individual inputs are not revealed. The results from our usability study show that a variety of client-side metrics can be collected at the aggregate level to guide future usability improvements. This also demonstrates the relative ease with which this can be done: no new cryptographic tools or protocols (beyond those already necessary for the application) were built for this usability study.

Data sanitization, error detection, and error recovery cannot be achieved through manual inspection of secret-shared input data, and one single erroneous input may corrupt the entire result of an aggregate computation. Recovery from such errors is difficult and at worst impossible: it may require the execution of expensive MPC recovery protocols over the secret-shared data. This is further complicated by the fact that contributors must consent to the execution of such additional protocols. Thus, it is important that extensive error checking be performed on the client side within the application before data sets are contributed. Even when the contributed data set consists of analytics or usability metrics, such cleaning cannot be done ad hoc after the computation and must be encoded into the MPC protocol before deployment occurs.

The prohibitive cost of sanitizing the data after submission also makes it critical that contributors can resubmit (or withdraw) their data. In our last deployment, we received an email from a participant inquiring if resubmission is possible, as they had submitted random data to try the application out. Without resubmission, the output would have been corrupted in its entirety.

Finally, we have found that our experience echoes and confirms thoughts expressed by other researchers in the community on the development of real-world MPC applications [53]: "choose an application, starting from a very real business need, and build the solution from the problem itself choosing the right tools, tuning protocol ideas into a reasonable solution, balancing security and privacy needs vs. other constraints: legal, system setting, etc."

## Acknowledgments

This work is partially supported by the NSF (under Grants #1430145, #1414119, #1718135, and #1739000) and the Honda Research Institutes. We are grateful to the Boston Women's Workforce Council and the Greater Boston Chamber of Commerce for their continued partnership in deploying MPC. We would like to acknowledge Rose Kelly for her work on the heuristics evaluation. We thank our shepherd, Nina Taft, for the exceptional care and thought that went into her advice, and our reviewers for their detailed and insightful feedback.

## Availability

The entire codebase is open-source and available at: https://github.com/multiparty/web-mpc

## References


[1] U.S. Equal Employment Opportunity Commission EEO-1 Survey. https://www.eeoc.gov/employers/eeo1survey/index.cfm. [Accessed: March 7, 2017].

[2] VIFF, the Virtual Ideal Functionality Framework. http://viff.dk/. [Accessed: August 15, 2015].

[3] Istemi Ekin Akkus, Ruichuan Chen, Michaela Hardt, Paul Francis, and Johannes Gehrke. Non-tracking web analytics. In *Proceedings of the 2012 ACM Conference on Computer and Communications Security*, CCS '12, pages 687–698, New York, NY, USA, 2012. ACM.

[4] Richard Atterer, Monika Wnuk, and Albrecht Schmidt. Knowing the user's every move: User activity tracking for website usability evaluation and implicit interaction. In *Proceedings of the 15th International Conference on World Wide Web*, WWW '06, pages 203–212, New York, NY, USA, 2006. ACM.

[5] Aaron Bangor, Philip Kortum, and James Miller. Determining what individual sus scores mean: Adding an adjective rating scale. *J. Usability Studies*, 4(3):114–123, May 2009.

[6] Rich Barlow. Computational Thinking Breaks a Logjam. http://www.bu.edu/today/2015/computational-thinking-breaks-a-logjam/. [Accessed: August 15, 2015].

[7] Johes Bater, Gregory Elliott, Craig Eggen, Satyender Goel, Abel N. Kho, and Jennie Rogers. SMCQL: secure query processing for private data networks. *PVLDB*, 10(6):673–684, 2017.

[8] Mihir Bellare, Viet Tung Hoang, Sriram Keelveedhi, and Phillip Rogaway. Efficient garbling from a fixed-key blockcipher. In *2013 IEEE Symposium on Security and Privacy, SP 2013, Berkeley, CA, USA, May 19-22, 2013*, pages 478–492, 2013.

[9] Mihir Bellare, Viet Tung Hoang, and Phillip Rogaway. Foundations of garbled circuits. In *the ACM Conference on Computer and Communications Security, CCS'12, Raleigh, NC, USA, October 16-18, 2012*, pages 784–796, 2012.

[10] Assaf Ben-David, Noam Nisan, and Benny Pinkas. Fairplaymp: a system for secure multi-party computation. In *Proceedings of the 15th ACM conference on Computer and communications security*, pages 257–266. ACM, 2008.

[11] Dan Bogdanov, Liina Kamm, Baldur Kubo, Reimo Rebane, Ville Sokk, and Riivo Talviste. Students and Taxes: a Privacy-Preserving Study Using Secure Computation. *PoPETs*, 2016(3):117?135, 2016.

[12] Dan Bogdanov, Sven Laur, and Jan Willemson. Sharemind: A Framework for Fast Privacy-Preserving Computations. In Sushil Jajodia and Javier Lopez, editors, *Proceedings of the 13th European Symposium on Research in Computer Security - ESORICS'08*, volume 5283 of *Lecture Notes in Computer Science*, pages 192–206. Springer Berlin / Heidelberg, 2008.

[13] Peter Bogetoft, Dan Lund Christensen, Ivan Damgård, Martin Geisler, Thomas Jakobsen, Mikkel Krøigaard, Janus Dam Nielsen, Jesper Buus Nielsen, Kurt Nielsen, Jakob Pagter, Michael Schwartzbach, and Tomas Toft. Financial cryptography and data security. chapter Secure Multiparty Computation Goes Live, pages 325–343. Springer-Verlag, Berlin, Heidelberg, 2009.

[14] Boston Women's Workforce Council. Boston women's workforce council report 2016. January 2017.

[15] Boston Women's Workforce Council. Boston women's workforce council report 2017. January 2018.

[16] John Brooke. Sus: a retrospective. *Journal of usability studies*, 8(2):29–40, 2013.

[17] Martin Burkhart, Mario Strasser, Dilip Many, and Xenofontas Dimitropoulos. Sepia: Privacy-preserving aggregation of multi-domain network events and statistics. In *USENIX SECURITY SYMPOSIUM*. USENIX, 2010.

[18] Chris Clifton, Murat Kantarcioglu, Jaideep Vaidya, Xiaodong Lin, and Michael Y. Zhu. Tools for privacy preserving distributed data mining. *SIGKDD Explor. Newsl.*, 4(2):28–34, December 2002.

[19] Boston Women's Workforce Council. What We Do. accessed 10/10/2018.

[20] Boston Women's Workforce Council. Boston women's workforce council 2017, Januaray 2018.

[21] L.F. Cranor and S. Garfinkel. *Security and Usability: Designing Secure Systems that People Can Use*. O'Reilly Media, 2005.

[22] Ivan Damgård, Kasper Damgård, Kurt Nielsen, Peter Sebastian Nordholt, and Tomas Toft. Confidential benchmarking based on multiparty computation. Cryptology ePrint Archive, Report 2015/1006, 2015. http://eprint.iacr.org/.

[23] Daniel Demmler, Thomas Schneider, and Michael Zohner. Aby-a framework for efficient mixed-protocol secure two-party computation. In *NDSS*, 2015.

[24] Yael Ejgenberg, Moriya Farbstein, Meital Levy, and Yehuda Lindell. Scapi: The secure computation application programming interface. iacr cryptology eprint archive, 2012.

[25] Oded Goldreich. *The Foundations of Cryptography - Volume 2, Basic Applications*. Cambridge University Press, 2004.

[26] Adam Groce, Alex Ledger, Alex J. Malozemoff, and Arkady Yerukhimovich. Compgc: Efficient offline/online semi-honest two-party computation. *IACR Cryptology ePrint Archive*, 2016:458, 2016.

[27] Marcella Hastings, Brett Hemenway, Daniel Noble, and Steve Zdancewic. Sok: General purpose compilers for secure multi-party computation. In *SoK: General Purpose Compilers for Secure Multi-Party Computation*, page 0. IEEE, 2019.

[28] Wilko Henecka, Stefan Kögl, Ahmad-Reza Sadeghi, Thomas Schneider, and Immo Wehrenberg. TASTY: tool for automating secure two-party computations. In *Proceedings of the 17th ACM Conference on Computer and Communications Security, CCS 2010, Chicago, Illinois, USA, October 4-8, 2010*, pages 451–462, 2010.

[29] Ayman Jarrous and Benny Pinkas. Canon-mpc, a system for casual non-interactive secure multi-party computation using native client. In *Proceedings of the 12th ACM Workshop on Workshop on Privacy in the Electronic Society*, WPES '13, pages 155–166, New York, NY, USA, 2013. ACM.

[30] Ronald Kainda, Ivan Flechais, and AW Roscoe. Security and usability: Analysis and evaluation. In *2010 International Conference on Availability, Reliability and Security*, pages 275–282. IEEE, 2010.

[31] William H. Kruskal and W. Allen Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260):583–621, 1952.

[32] Andrei Lapets, Nikolaj Volgushev, Azer Bestavros, Frederick Jansen, and Mayank Varia. Secure multi-party computation for analytics deployed as a lightweight web application. Technical report, Computer Science Department, Boston University, 2016.

[33] John Launchbury, Iavor S. Diatchki, Thomas DuBuisson, and Andy Adams-Moran. Efficient lookup-table protocol in secure multiparty computation. In *Proceedings of the 17th ACM SIGPLAN International Conference on Functional Programming*, ICFP '12, pages 189–200, New York, NY, USA, 2012. ACM.

[34] Kevin E. Levay, Jeremy Freese, and James N. Druckman. The demographic and political composition of mechanical turk samples. *SAGE Open*, 6(1):2158244016636433, 2016.

[35] Joanne Lipman. Let's Expose the Gender Pay Gap. http://www.nytimes.com/2015/08/13/opinion/lets-expose-the-gender-pay-gap.html. [Accessed: August 15, 2015].

[36] Chang Liu, Xiao Shaun Wang, Kartik Nayak, Yan Huang, and Elaine Shi. Oblivm: A programming framework for secure computation. In *IEEE S & P*, 2015.

[37] Alfred J. Menezes, Scott A. Vanstone, and Paul C. Van Oorschot. *Handbook of Applied Cryptography*. CRC Press, Inc., Boca Raton, FL, USA, 1st edition, 1996.

[38] Kartik Nayak, Xiao Shaun Wang, Stratis Ioannidis, Udi Weinsberg, Nina Taft, and Elaine Shi. Graphsc: Parallel secure computation made easy. In *2015 IEEE Symposium on Security and Privacy, SP 2015, San Jose, CA, USA, May 17-21, 2015*, pages 377–394, 2015.

[39] Jakob Nielsen. *Usability Engineering*. Morgan Kaufmann, San Francisco, CA, 1993.

[40] Jakob Nielsen and Robert L. Mack. *Usability Inspection Methods*. John Wiley & Sons, Inc., New York, 1994.

[41] Greater Boston Chamber of Commerce. Gbcc launches pacesetters initiative aimed at uniting the business community's response to economic inclusion, January 2018. Retrieved February 26, 2019 from http://bostonchamber.com/about-us/media-center/gbcc-launches-pacesetters-initiative.

[42] Greater Boston Chamber of Commerce. Pacesetters initiative, January 2018. Retrieved February 26, 2019 from http://bostonchamber.com/programs-events/pacesetters.

[43] Do Le Quoc, Martin Beck, Pramod Bhatotia, Ruichuan Chen, Christof Fetzer, and Thorsten Strufe. Privapprox: Privacy-preserving stream analytics. In *2017 USENIX Annual Technical Conference (USENIX ATC 17)*, pages 659–672, Santa Clara, CA, 2017. USENIX Association.

[44] Aseem Rastogi, Matthew A. Hammer, and Michael Hicks. Wysteria: A programming language for generic, mixed-mode multiparty computations. In *Proceedings of the 2014 IEEE Symposium on Security and Privacy*, SP '14, pages 655–670, Washington, DC, USA, 2014. IEEE Computer Society.

[45] Scott Ruoti, Jeff Andersen, Daniel Zappala, and Kent E. Seamons. Why johnny still, still can't encrypt: Evaluating the usability of a modern PGP client. *CoRR*, abs/1510.08555, 2015.

[46] Martina Angela Sasse, Sacha Brostoff, and Dirk Weirich. Transforming the 'weakest link'—a human/computer interaction approach to usable and effective security. *BT technology journal*, 19(3):122–131, 2001.

[47] Berry Schoenmakers. Mpyc: Secure multiparty computation in python, 2018. https://www.win.tue.nl/berry/mpyc/.

[48] Axel Schroepfer and Florian Kerschbaum. Demo: Secure computation in javascript. In *Proceedings of the 18th ACM Conference on Computer and Communications Security*, CCS '11, pages 849–852, New York, NY, USA, 2011. ACM.

[49] Adi Shamir. How to share a secret. *Commun. ACM*, 22(11):612–613, November 1979.

[50] Nikolaj Volgushev, Malte Schwarzkopf, Ben Getchell, Mayank Varia, Andrei Lapets, and Azer Bestavros. Conclave: secure multi-party computation on big data. 2019.

[51] Alma Whitten and J. D. Tygar. Why johnny can't encrypt: A usability evaluation of pgp 5.0. In *Proceedings of the 8th Conference on USENIX Security Symposium - Volume 8*, SSYM'99, pages 14–14, Berkeley, CA, USA, 1999. USENIX Association.

[52] Ka-Ping Yee. Aligning security and usability. *IEEE Security & Privacy*, 2(5):48–55, 2004.

[53] Moti Yung. From mental poker to core business: Why and how to deploy secure computation protocols? In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, CCS '15, pages 1–2, New York, NY, USA, 2015. ACM.

## A Heuristic Evaluation

Two evaluators, both of whom were familiar with general human factors and usability principles, conducted the heuristic evaluation. The evaluators were familiar with the web application, but had not been involved in its development or deployment. Each independently examined the web application and listed general usability concerns. Next, they emulated users and attempted to submit data. While emulating the user tasks, they documented any issues they found. Then, they independently categorized the issues into the ten heuristic categories listed in Tables 4 and 5 [39, 40].

After each issue was placed into a category, each evaluator gave each issue a severity rating according to the frequency, impact, and persistence of the issue. The severity ratings ranged from 1 to 5, with 1 representing the lowest severity and 5 representing the highest severity. Next, the evaluators compiled a list of potential solutions. Lastly, the evaluators met with the web application designers to discuss the feasibility of implementing the solutions.

## B User Interface Redesign

The redesign divides the application into four distinct steps, each visually and logically separated by a *card layout*. Cards adhere to a pattern of *help text*, followed by a separator and the *action item*. The first card contains three input elements: session ID, participation ID, and spreadsheet (cf. Figure 7). Both the session and participation ID are validated when the user shifts the application focus from the input box. The final input element, the spreadsheet, triggers a browser-based parser for Excel files. This allows participants to simply drag-and-drop the pre-filled template into the web application, and reduce the possibility of copy/paste mistakes.

The second card contains the various tables for data entry, as displayed in Figure 8. It shows only the help text by default. To encourage participants to upload data rather than entering it manually, the card only expands after the spreadsheet has been provided. The tables were originally split by gender, and contained entries for financial information, length of service, and employee counts. User testing showed this caused confusion, so our new interface reverses the grouping. This resolves issues.

Validation of entries was enhanced for the 2017 collection. First, "semantic validation" across tables was added: if the employee count for a category in the first table is greater than zero, we also expect a matching salary and length of service greater than zero for that field in other tables. Clicking on any red cells shows a pop-up describing the problem. This resolves issues. The same validation is also present in the Excel spreadsheet, so participants get the chance to address any issues early on.

With the third card we introduce a new feature to collect anonymous user feedback about the data collection process. The responses will inform improvements for future data sessions. Answering questions is mandatory in this iteration, and unanswered questions turn red when attempting to submit.

The fourth and final card handles the actual data submission. In this card, shown in Figure 9, we show a summary of the employee count entered, a list of all errors messages that prevent a successful submission, and the history of both successful and failed submissions. A pop-up appears on submission to highlight the submission outcome.

## C Secure Aggregation Protocol

In this section, we describe in full our protocol for secure aggregation within a finite additive group $G$ such as $\mathbb{Z}/2^{64}\mathbb{Z}$. Let $\Sigma = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ be a public-key encryption scheme that provides IND-CPA security; concretely, we use RSA in our implementation [37]. We restrict our attention to the case with a single analyst (as with the pay equity scenario). A single execution or *session* of the protocol proceeds in the following way:

Table 4: Listing of usability issues and average severity ratings categorized by heuristics 1 through 6; for the severity ratings, 1 is the lowest severity and 5 is the highest severity. Issues not fixed in the redesign are marked.

| # | Usability Issue | Avg. | Fixed |
|---|---|---|---|
| **1** | **Visibility of system status: the web application should always keep users informed, through appropriate feedback within reasonable time.** | | |
| 1.1 | There is no indication as how much of the table has been or is yet to be completed. | 4.5 | |
| 1.2 | There is no indication as to whether the session key is valid. | 3.5 | |
| 1.3 | There is no indication as to whether the email address is valid. | 1.5 | |
| 1.4 | After submission, user sees messages saying "loading" and then a confirmation window, which is confusing. | 3 | |
| 1.5 | After submission, there is no information indicating that data can be resubmitted. | 3.5 | |
| 1.6 | There is no email confirmation indicating that data was submitted. | 2 | |
| **2** | **Match between system and the real world: the web application should speak the user's language and present information in a logical order.** | | |
| 2.1 | The column and row headings do not use real-world terms that Human Resources (HR) uses, e.g., sum instead of total, workforce instead of employees, and mos. instead of months. | 4.5 | No |
| 2.2 | The tables are separated by gender, irrespective of whether HR data is usually separate by gender, e.g., if it is separated by ethnicity, it will be difficult for them to enter data separated by gender. | 4.5 | No |
| 2.3 | The table require a summation instead of an average, irrespective of whether HR data is given via averages or summations. | 4 | No |
| 2.4 | The columns requiring summary data (i.e., sum) are visually the same and not separated from data on raw numbers or monetary values. | 2 | |
| 2.5 | The sum cells require one to calculate totals by hand. | 3.5 | No |
| 2.6 | When you drag to select the same value for multiple cells, the cells are highlighted in red, implying an error. | 3.5 | |
| **3** | **User control and freedom: users will make mistakes and should be able to fix their errors easily. The web application should support undo, redo, and process cancellation.** | | |
| 3.1 | A cell is highlighted in red if a user clicks there, does not input a number, then clicks somewhere else. | 5 | |
| 3.2 | Ctrl + Z (undo) is functional, but it always results in the previous cell being highlighted in red. | 4 | |
| 3.3 | The meaning of the red cell is unclear. | 5 | |
| 3.4 | Decimal points are not allowed in any cell. | 4 | |
| **4** | **Consistency and standards: users should not have to wonder whether different words, situations, or actions mean the same thing. The web application should follow platform conventions.** | | |
| 4.1 | The terms #, $, and mos. are used in the row headings, but not the column headings. | 3.5 | No |
| 4.2 | The difference between multiple employee groups is unclear, e.g., executive versus mid-level. | 5 | No |
| 4.3 | There is no option for "other" employee, i.e., if they don't fall into one of the employee groups. | 5 | No |
| 4.4 | While there is an option for 2+ races, not including Hispanic/Latino, there is not an option for 2+ races, including Hispanic/Latino. | 5 | No |
| 4.5 | Some employee types end with the word worker, but others do not. | 2.5 | No |
| **5** | **Error prevention: Reduce errors by reconfirming actions before they are carried out.** No errors were detected. | | |
| **6** | **The user should not have to remember information from one part of the dialogue to another. Instructions for use of the web application should be visible or easily retrievable whenever appropriate.** | | |
| 6.1 | There is no objective or set of instructions indicating what and where information is to be entered as well as where users can find the session key or appropriate email address. | 5 | |
| 6.2 | There are no definitions of the terms, e.g., executive, mid-level, and annual compensation | 5 | No |

Table 5: Listing of usability issues and average severity ratings categorized by heuristics 7 through 10; for the severity ratings, 1 is the lowest severity and 5 is the highest severity. Issues not fixed in the redesign are marked.

| | | | |
|---|---|---|---|
| **7** | **Flexibility and efficiency of use: The web application should cater to both inexperienced and experienced users, allow users to tailor frequent actions, and provide defaults.** | | |
| 7.1 | Though copy and paste works, if an empty cell is copied, the pasted cell will be highlighted in red, indicating that the copy and paste procedure did not work. | 4 | No |
| 7.2 | There is no way to enter functions into the cells, e.g., C2 = A2 + B2. | 2 | |
| 7.3 | You can only drag cells to copy values either horizontally or vertically, not both. | 2.5 | |
| **8** | **Aesthetic and minimalist design: dialogues should not contain information which is irrelevant or rarely needed.** | | |
| 8.1 | Contrast between column and row fillings and text may be inadequate, i.e., black text on grey background may not be visible for some users. | 4 | |
| 8.2 | Red cells are inappropriate for those who are color blind, which is 8 percent of all males. | 5 | |
| **9** | **Help users recognize, diagnose, and recover from errors: error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.** | | |
| 9.1 | There are no messages associated with the red cells. | 5 | |
| 9.2 | There are no messages associated with the grey cells. | 3.5 | |
| 9.3 | There is not a list of errors near the submit button that would indicate what needs to be fixed before submission is possible. | 4 | |
| **10** | **Help documentation: Documentation should be easy to search, be focused on the user's task, list concrete steps to be carried out, and be minimalist.** | | |
| 10.1 | There is no help page, documentation, or instructions. | 5 | |

1. The analyst initiates the process by generating a secret and public RSA key pair $(s, p)$ and a unique session identifier $\mathsf{id} \in \mathbb{N}$, submitting $p$ to the service provider, and sending $\mathsf{id}$ to all the contributors;[4]

2. Each of the $n$ contributors possesses a secret *data* value $d_i \in G$ and does the following at least once[5]:

   (a) Generate a secret *random mask* $m_i \in G$ and calculate the *masked data* $r_i = d_i + m_i$,

   (b) Receive $p$ from the service provider.

   (c) Send $r_i$ and $c_i = \mathrm{Enc}_p(m_i)$ to the service provider.

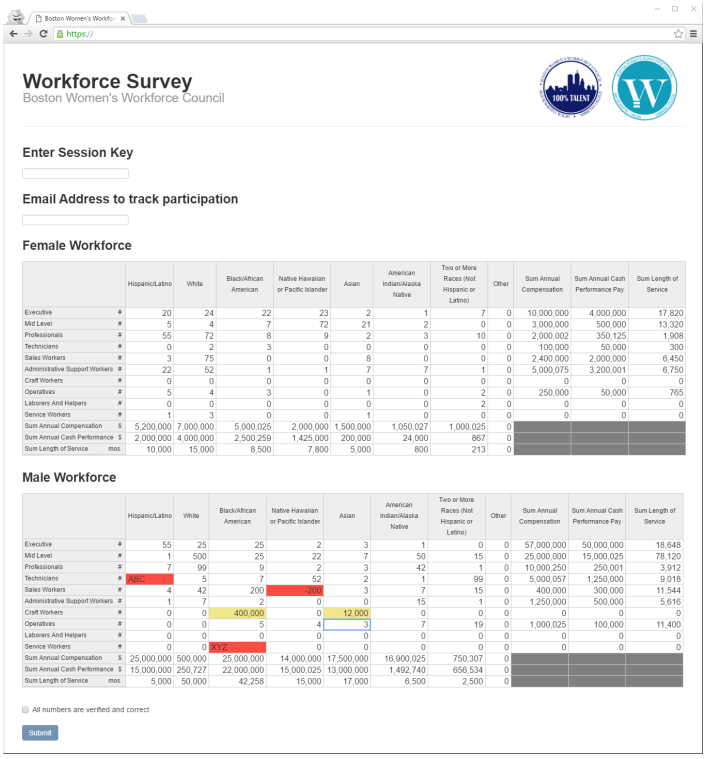
3. The service provider computes the sum of the masked data values to obtain the aggregate masked data quantity $R = \sum_{i=1}^{n} r_i$;

4. The analyst then retrieves $R$ and all the $c_1, \ldots, c_n$ from the service provider, computes $m_i = \mathrm{Dec}_s(c_i)$ for all $i$, computes $M = \sum_{i=1}^{n} m_i$, and obtains the final result $R - M = \sum_{i=1}^{n} d_i$. No other party receives any output.

Figure 1 illustrates an example deployment of the protocol with two contributors. Intuitively, this protocol is secure because the service provider's view of the random masks is protected using the analyst's public key, and the analyst never sees the individual masked data values unless it violates its promise not to collude with the service provider.

[4] The session identifier is only to allow distinct sessions, but it can serve another purpose: if no malicious agent possesses the session identifier, any data submitted by malicious agents will be ignored during the computation of the result.

[5] Each contributor can perform step (2) as many times as they wish before step (3) occurs; the operation they perform is idempotent if they always submit the same data.



Workforce Survey
Boston Women's Workforce Council

Enter Session Key

Email Address to track participation

Female Workforce

| | | Hispanic/Latino | White | Black/African American | Native Hawaiian or Pacific Islander | Asian | American Indian/Alaska Native | Two or More Races (Not Hispanic or Latino) | Other | Sum Annual Compensation | Sum Annual Cash Performance Pay | Sum Length of Service |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Executive | # | 20 | 24 | 22 | 23 | 2 | 1 | 7 | 0 | 10,000,000 | 4,000,000 | 17,820 |
| Mid Level | # | 5 | 4 | 7 | 72 | 21 | 2 | 0 | 0 | 3,000,000 | 500,000 | 13,320 |
| Professionals | # | 55 | 72 | 8 | 9 | 2 | 3 | 10 | 0 | 2,000,002 | 350,125 | 1,908 |
| Technicians | # | 0 | 2 | 3 | 0 | 0 | 0 | 0 | 0 | 100,000 | 50,000 | 300 |
| Sales Workers | # | 3 | 75 | 0 | 0 | 8 | 0 | 0 | 0 | 2,400,000 | 2,000,000 | 6,450 |
| Administrative Support Workers | # | 22 | 52 | 1 | 1 | 7 | 7 | 1 | 0 | 5,000,075 | 3,200,001 | 6,750 |
| Craft Workers | # | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Operatives | # | 5 | 4 | 3 | 0 | 1 | 0 | 2 | 0 | 250,000 | 50,000 | 765 |
| Laborers And Helpers | # | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 |
| Service Workers | # | 1 | 3 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Sum Annual Compensation | $ | 5,200,000 | 7,000,000 | 5,000,025 | 2,000,000 | 1,500,000 | 1,050,027 | 1,000,025 | 0 | | | |
| Sum Annual Cash Performance | $ | 2,000,000 | 4,000,000 | 2,500,259 | 1,425,000 | 200,000 | 24,000 | 867 | 0 | | | |
| Sum Length of Service | mos. | 10,000 | 15,000 | 8,500 | 7,800 | 5,000 | 800 | 213 | 0 | | | |

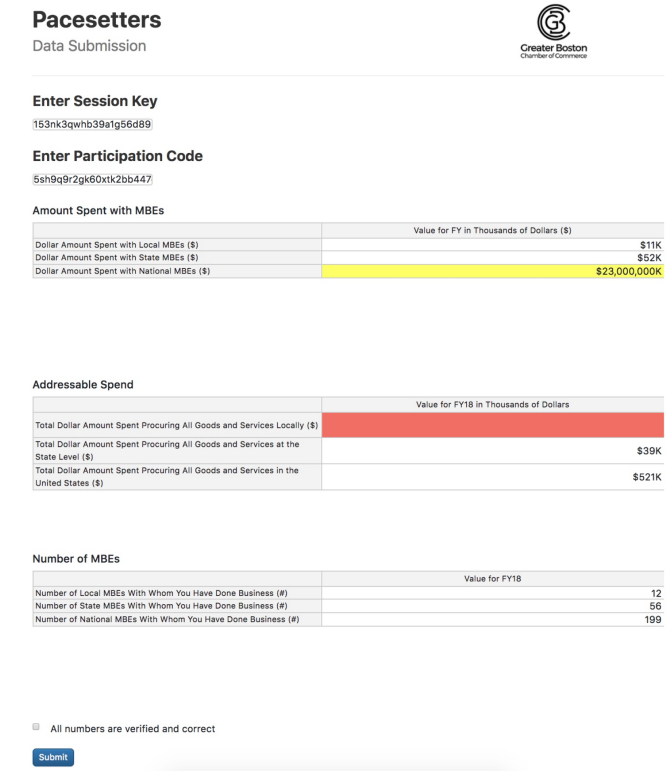
Male Workforce

| | | Hispanic/Latino | White | Black/African American | Native Hawaiian or Pacific Islander | Asian | American Indian/Alaska Native | Two or More Races (Not Hispanic or Latino) | Other | Sum Annual Compensation | Sum Annual Cash Performance Pay | Sum Length of Service |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Executive | # | 55 | 25 | 25 | 2 | 3 | 1 | 0 | 0 | 57,000,000 | 50,000,000 | 18,648 |
| Mid Level | # | 1 | 500 | 25 | 22 | 7 | 50 | 15 | 0 | 25,000,000 | 15,000,025 | 78,120 |
| Professionals | # | 7 | 99 | 9 | 2 | 3 | 42 | 1 | 0 | 10,000,250 | 250,001 | 3,912 |
| Technicians | # | ABC | 5 | 7 | 52 | 2 | 1 | 99 | 0 | 5,000,057 | 1,250,000 | 9,018 |
| Sales Workers | # | 4 | 42 | 200 | -200 | 3 | 7 | 15 | 0 | 400,000 | 300,000 | 11,544 |
| Administrative Support Workers | # | 1 | 7 | 2 | 0 | 0 | 15 | 1 | 0 | 1,250,000 | 500,000 | 5,616 |
| Craft Workers | # | 0 | 0 | 400,000 | 0 | 12,000 | 0 | 0 | 0 | 0 | 0 | 0 |
| Operatives | # | 0 | 0 | 5 | 4 | 3 | 7 | 19 | 0 | 1,000,025 | 100,000 | 11,400 |
| Laborers And Helpers | # | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Service Workers | # | 0 | 0 | XYZ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Sum Annual Compensation | $ | 25,000,000 | 500,000 | 25,000,000 | 14,000,000 | 17,500,000 | 16,900,025 | 750,307 | 0 | | | |
| Sum Annual Cash Performance | $ | 15,000,000 | 250,727 | 22,000,000 | 15,000,025 | 13,000,000 | 1,492,740 | 656,534 | 0 | | | |
| Sum Length of Service | mos. | 5,000 | 50,000 | 42,258 | 15,000 | 17,000 | 6,500 | 2,500 | 0 | | | |

All numbers are verified and correct

Submit

Figure 4: The original contributor web interface used for the BWWC study at https://100talent.org as it appears within a web browser with some user errors highlighted.



Pacesetters
Data Submission

Greater Boston Chamber of Commerce

Enter Session Key

153nk3qwhb39a1g56d89

Enter Participation Code

5sh9q9r2gk60xtk2bb447

Amount Spent with MBEs

| | Value for FY in Thousands of Dollars ($) |
|---|---|
| Dollar Amount Spent with Local MBEs ($) | $11K |
| Dollar Amount Spent with State MBEs ($) | $52K |
| Dollar Amount Spent with National MBEs ($) | $23,000,000K |

Addressable Spend

| | Value for FY18 in Thousands of Dollars |
|---|---|
| Total Dollar Amount Spent Procuring All Goods and Services Locally ($) | |
| Total Dollar Amount Spent Procuring All Goods and Services at the State Level ($) | $39K |
| Total Dollar Amount Spent Procuring All Goods and Services in the United States ($) | $521K |

Number of MBEs

| | Value for FY18 |
|---|---|
| Number of Local MBEs With Whom You Have Done Business (#) | 12 |
| Number of State MBEs With Whom You Have Done Business (#) | 56 |
| Number of National MBEs With Whom You Have Done Business (#) | 199 |

All numbers are verified and correct

Submit

Figure 5: The original interface modified for the Pacesetters Initiative. This was used by Group 1 in the usability study.

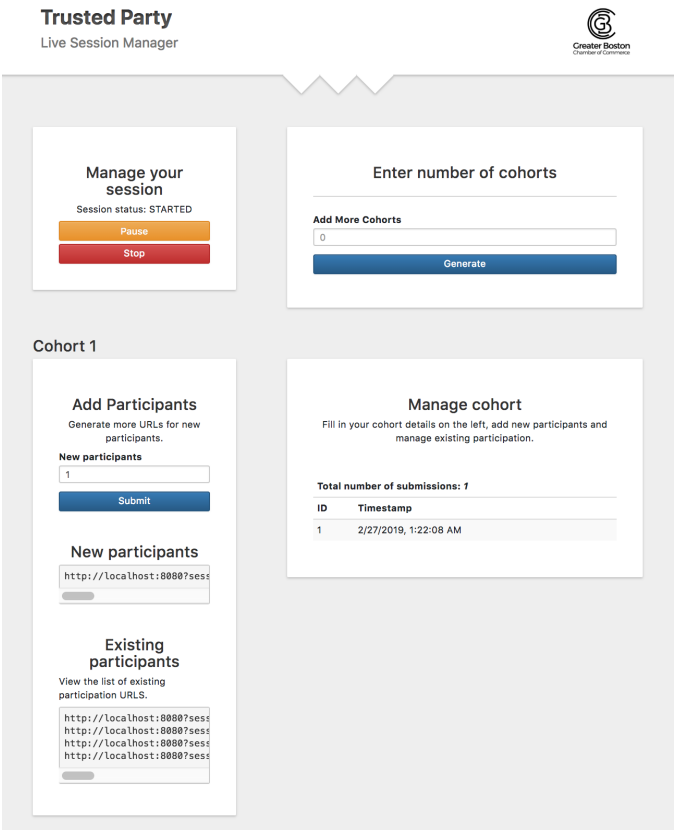



Figure 6: Page for analyst to create cohorts, generation participation codes, and track anonymous submission history

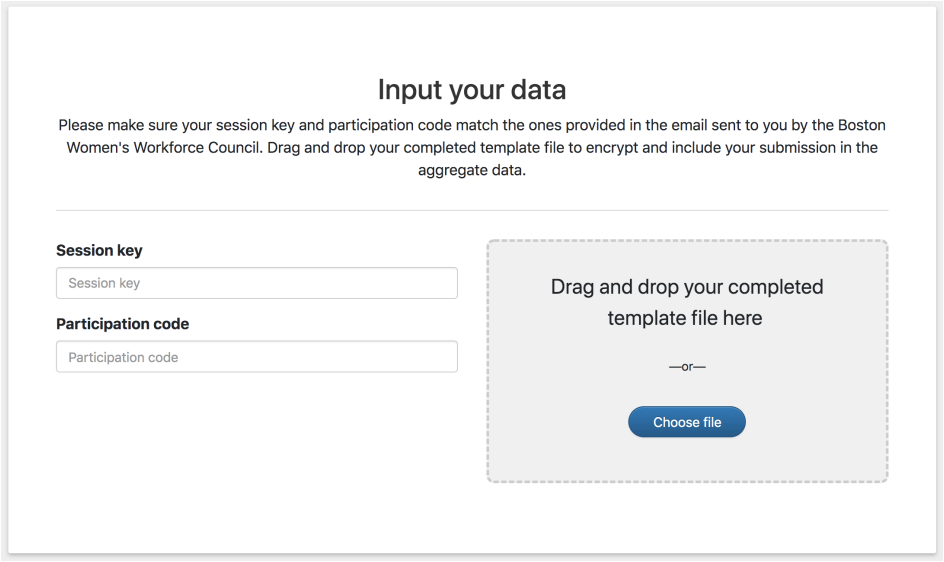



Figure 7: First card of the 2017 interface. To encourage drag-and-drop upload from an Excel file, this is the only card shown by default.

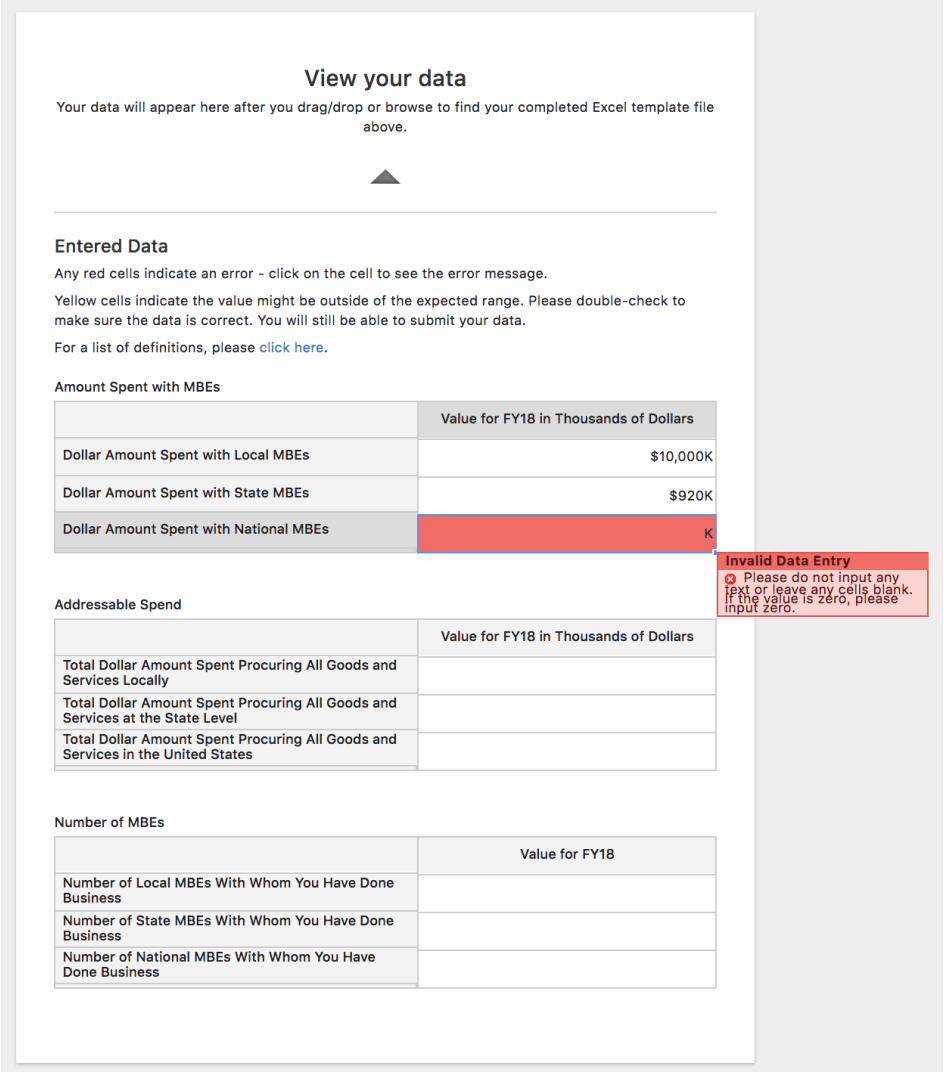



Figure 8: Submission card within the V2 interface. This example displays an empty cell and the corresponding tooltip.

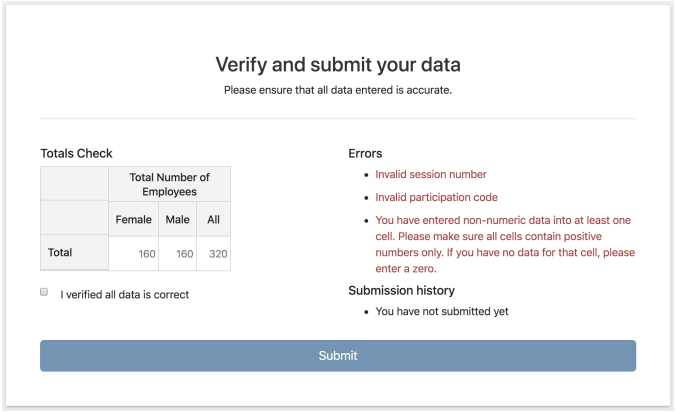



Figure 9: Final card within the V2 at https://100talent.org. In this example, the verification check has failed. Text boxes are still highlighted just as they were in the old interface (cf. Figure 4). Now, the full list of errors is co-located in this card.

# Certified Phishing: Taking a Look at Public Key Certificates of Phishing Websites

Vincent Drury
*Department of Computer Science*
*RWTH Aachen University*
*drury@itsec.rwth-aachen.de*

Ulrike Meyer
*Department of Computer Science*
*RWTH Aachen University*
*meyer@itsec.rwth-aachen.de*

## Abstract

The share of phishing websites using HTTPS has been constantly increasing over the last years. As a consequence, the simple user advice to check whether a website is HTTPS-protected is no longer effective against phishing. At the same time, the use of certificates in the context of phishing raises the question if the information contained in them could be used to detect phishing websites. In this paper we take a first step towards answering this question. To this end, we analyze almost 10 000 valid certificates queried from phishing websites and compare them to almost 40 000 certificates collected from benign sites. Our analysis shows that it is generally impossible to differentiate between benign sites and phishing sites based on the content of their certificates alone. However, we present empirical evidence that current phishing websites for popular targets do typically not replicate the issuer and subject information.

## 1 Introduction

Phishing is still an important and direct risk to many Internet users. The Anti-Phishing Working Group (APWG) recorded more than 50 000 unique phishing websites in September 2018, a number that has been relatively stable for the last year [19]. These websites follow the general trend of the Web [17] in that they are steadily adopting the usage of HTTPS: 49.4 % of the phishing sites were using SSL/TLS in the third quarter of 2018, up from less than 5 % in 2016. This rapid development, in conjunction with the availability of easy-to-obtain certificates, has already led to changes in browser design (e.g., Google Chrome aims to remove some positive security indicators [6]) and will render the general advice to "look for the lock icon" to detect phishing less and less effective.

There are mainly two lines of work aiming at protecting against phishing from two different directions: technical solutions and user educational approaches. These two lines complement one another rather than competing against each other as the former addresses the technical component while the latter addresses the social engineering aspect of phishing. The technical approaches include reactive measures, like blacklists, as well as preventive measures, like heuristic and machine learning-assisted detection approaches. Educational endeavors on the other hand focus on users, and try to improve their phishing detection and prevention abilities as soon as technical solutions fail. Previous educational efforts (e.g., [3, 22, 33]) mainly focus on noticing the absence of HTTPS usage or detecting suspicious URLs as indicators for phishing. With the rise of HTTPS-hosted phishing sites, the usage of HTTPS is no longer a strong indication for a benign site. However, certificates used by phishing sites are now a new potential source of information that might be useful in the context of automatic or user-based phishing detection.

Whether or not certificate information can be used in this context depends on the answers to the following open issues: First, it depends on whether there are differences between certificates of benign websites and phishing websites in the first place. Second, it depends on whether these differences are robust, i.e., whether it is safe to assume that they persist even if adversaries try to actively reduce these differences. Third, even if there are robust differences, it remains an open question, whether these differences can effectively be used to discriminate between phishing and benign certificates as part of a technical solution and/or whether these differences can be exposed to a user in a way such that they increase the user's ability to detect a phishing website.

In this paper, we focus on addressing the first of these issues and briefly touch the second one. The third of the above issues is not addressed in this paper.


*USENIX Symposium on Usable Privacy and Security (SOUPS)* 2019.
August 11–13, 2019, Santa Clara, CA, USA.

More specifically, we empirically investigate the following two research questions:

1. Are there general differences between the certificates of phishing websites compared to those of benign websites and if so which ones?

2. Are there differences between the certificate of a phishing website and the certificate of the corresponding targeted benign website and if so which ones?

To answer these questions, we collect and compare 9 479 certificates from 31 264 phishing websites and 39 478 certificates from 50 000 benign websites.

We find no obvious differences between phishing and benign websites in general. However, we find that the phishing certificates of the 15 most popular phishing target's websites do currently differ from their benign counterparts in particular with respect to the issuer and subject information provided in the certificate.

We also identify the threat of hosting services, that make it easy for attackers to present a valid certificate that looks very similar to the target's certificate, in particular if the target is the hoster itself. Even users that have knowledge of URLs, may fall victim to such an attacker as the fake website is hosted on a legitimate domain.

The rest of this paper is structured as follows: The next section introduces some preliminary terms and concepts. Section 3 presents related work on the topics of phishing education. Section 4 details the certificate collection process, as well as the results of our analysis. Section 5 takes a look at the representation of certificate information in browsers. Finally, we conclude with a summary and future work.

## 2 Preliminaries

In this section we present some basic concepts and terms that will be used throughout this paper.

### 2.1 Website-based phishing attacks

In this paper, we look at the following website-based type of phishing attack, that we will generally refer to as phishing: The **attacker** clones the website of a **target** and sends a link to a user of the target, the **victim**. In particular, we do not restrict the transmission channel to email, other methods might also be possible. The victim then clicks on the link and opens the attacker's (fake) website and interacts with the phishing website as if it was the website of the target. This interaction will typically include entering the victim's username and password information into the fake website, thus enabling the attacker to impersonate the user to the target in the future. We are not concerned about specifics of the attack (e.g., circumvention of two-factor authentication, website cloning techniques, etc.), as long as a fake website is involved.

| Field | |
|---|---|
| Subject: | `Common Name` (`CN`) |
| | `Organization` (`O`) |
| | `Organizational Unit` (`OU`) |
| | `Locality` (`L`) |
| | `Country of Residence` (`C`) |
| | `Business Category` |
| Issuer: | `Common Name` (`CN`) |
| | `Organization` (`O`) |
| | `Country of Residence` (`C`) |
| | `Valid From` |
| | `Valid To` |
| Extensions: | `Subject Alternative Name` (`SAN`) |
| | `Certificate Policies` |

Table 1: Certificate fields and shortnames used in this paper.

### 2.2 HTTPS and public key certificates

HTTPS allows web servers to authenticate themselves to users based on X.509 certificates using TLS [31]. Such certificates are issued by Certification Authorities (CAs) and bind the public key of a web server to the identity of the web server. Table 1 illustrates how the identity of the web server and its issuing CA are represented in an X.509 certificate and what other fields included in a certificate are of interest in the context of this paper. Note that the web server's domain name is included in the certificate either in the subject `CN` field or as an entry in the `SAN` extension field.

The CAs used on the web today are ordered in a hierarchy, where CAs on higher levels issue certificates for CAs on lower levels, and the CAs at the lowest level issue certificates for the individual web servers. The certificates of the CAs at the highest level, the root CAs' certificates, are shipped in web browsers and are thus readily available on the client side. When a client connects to the web server with HTTPS, the web server presents a chain of certificates to the client and the client can validate the certificates in the chain, starting with checking that the last certificate in the chain was issued by one of the pre-established root CAs and thus obtaining a public key to check the next certificate in the chain. While certificates certainly help in validating public keys, the mere fact that a website is able to present a valid chain of certificates is not a guarantee that the website itself is trustworthy as CAs may follow different policies while issuing certificates. Thus, it is possible that a request for a certificate, e.g., for an intentionally misleading domain name, is indeed signed by a CA if the policy used by the CA to validate the identity of the requester is rather lax.

### 2.3 Types of Validation

There are several levels of vetting a CA can perform before signing a certificate for a subject, that can also influence the

amount of information included in the certificate. These validation types represent different levels of trust or effort by the CAs and are briefly introduced in the following. We use the CA/Browser Forum's (CAB) guidelines as reference for the different validation levels [4].

According to these guidelines, all CAs have to ensure certain qualities regardless of the type of validation, that include basic employee vetting as well as logging and auditing requirements. The CAs also have to ensure that all information that is included in a certificate was verified, taking reasonable steps to ensure correctness. In the context of phishing, it is worth mentioning that CAs are required to maintain a database of "high-risk" names, that are at risk for phishing or other fraudulent usage. This database has to be checked for each certificate that is issued, and if a high-risk name is found, additional scrutiny on the part of the CA is expected to make sure that the certificate is issued to a valid entity. There are, however, no specific requirements on how "high-risk" names are to be handled in the CAB documents.

#### Domain Validation

Domain Validation (DV) is the most basic form of validation. Here, the CA only checks that the Certificate Signing Request is valid and that the subject has control over the domain in question (indicated in the `CN` or `SAN` field of the certificate). This might include a challenge, e.g., setting a specific DNS entry or uploading a file with some predefined content. Since no further review is required to validate control over a domain, this process can be automated, e.g., as is the case with the CA "Let's Encrypt" [23].

#### Organization Validation

Certificates where the CA has asserted the validity of the subject's organization identity are called Organization Validated (OV) certificates. In the CAB documents [4], this requires more rigorous validation of the subject, beyond simple control of the domain. Verification of the organization identity means, that the issuing CA has to verify name and address of the organization entity, e.g., via consulting the government agency in the jurisdiction of the organization, or a site visit. Additionally, the CA has to verify the authenticity of the certificate applicant, e.g., via a reliable method of communication.

As a result of the organization validation, the CA is able to add organization information (i.e., the `O`, `OU`, `C`, `L` fields) to the certificate. They might also include the CAB policy ID for OV certificates (2.23.140.1.2.2) in the `Certificate Policies` field of the certificate, and must then include the subject field `O` as well as location information (i.e., country and state or province).

#### Extended Validation

The most thorough validation level is called Extended Validation (EV) and is used to validate the legal entity that controls a website [5]. Preventing phishing is explicitly mentioned as a secondary purpose, a consequence of the more reliable information included in the certificate. The main difference to OV certificates is, that the process for issuing EV certificates is defined in much more detail and adds some additional requirements. In theory, a CA could issue a non-EV certificate using the EV validation processes.

The documents include, among others, detailed requirements for certificate requests. For EV certificates, the certificate applicant has to name several contact people, who have to fulfill certain roles (certificate requester, certificate approver, certificate signer, applicant representative), all of which have to be authenticated by the CA. The CA also has to verify the organization's legal, physical and operational existence, verify the authority of all roles of requesters and ensure reliable means of communication in addition to verifying domain control. The guidelines also introduce additional constraints to EV certificates, including the prohibition of wildcard certificates. CAs will also have to look out for high risk certificates, that include websites with the risk of fraud (e.g., websites with an Internationalized Domain Name (IDN) [21] that looks similar to an existing business). An EV certificate must include several fields:

- The subject organization name.
- The subject business category (e.g., private organization or government entity).
- The subject jurisdiction of incorporation or registration.
- The subject registration number (identifying the subject in the registration agency at the jurisdiction of the subject).
- An EV policy identifier that confirms the CA's compliance to the CAB EV documents. This can be specific to each CA.

All steps of the issuance process have to be documented and reviewed before granting the certificate request, all discrepancies have to be resolved. In particular, no single person must be able to grant an EV certificate, corresponding control procedures have to be enforced. The CA's employees have to be trained and their trustworthiness ensured via background checks (e.g., employment history, professional references, education, criminal history).

A client, for example a browser, checking the validity of an EV certificate, has to check for the corresponding policies in the certificate and confirm, that the issuing CA is valid and known to adhere to the EV guidelines.

## 3 Related Work

In our work, we investigate at a large scale whether and if so how certificates of benign and phishing websites differ. As such, it is related to several fields of study. In the following, we will look at previous work in phishing user studies, educational and technical approaches to prevent phishing, as well as browser evaluations regarding the presentation of certificate information and validation level.

Phishing is an attack that directly targets users, such that several **user studies** have set out to understand how and why it works. Phishing has been shown to be effective, even if users are primed to look for it, and even if they have technical knowledge [2, 32]. In 2006, Dhamija et al. published the results of a user study to find out, why users are susceptible to phishing [9]. They find, that users generally focus on the body of a website to decide if it is legitimate, ignoring more robust indicators like the URL. More recently, these results are confirmed by Alsharnouby et al., who track eye movement of users and find that they do not spend much time looking at security indicators. Less than 15% of the time is spent looking at browser UI, only about 6% is spent focusing on "areas of interest" like the URL bar or lock icon. The authors do however find, that browser indicators can be very helpful: detection correlates to focus on browser UI [2]. Downs et al. look at detection strategies and their effectiveness, especially for phishing emails [13]. They find, that knowledge about cues and past experience is not enough to reliably detect phishing.

Consequently, to get users to behave more securely, researchers have designed and evaluated several **educational approaches**. For example, Kumaraguru et al. conducted a large-scale (>500 participants) study that shows, that phishing education using PhishGuru, an embedded training method, can be effective and even have long-term benefits [22]. To create an engaging and immersive experience, researchers have also created and evaluated learning games to teach phishing detection. Sheng et al., with Anti-Phishing Phil in 2007, identified the problem that the browser UI is largely ignored in favor of the website body and try to teach users to understand and focus on URLs [33]. Arachchilage et al. design and evaluate a mobile game to prevent phishing [3]. Similarly to Phishing Phil, the game focuses on URLs. They show, that participants were motivated and improved their test scores after playing the game. These games mainly focus on URLs as indicators for phishing while, to the best of our knowledge, certificates have not been evaluated for user education so far.

User education as an approach has been shown to be somewhat successful, but no "perfect" results have been achieved. This leads to a different research direction, that focuses on automated **technical approaches** to phishing prevention to support and complement user efforts. A widely represented approach are blacklists, that maintain lists of known phishing websites and prevent users from opening them. These lists can be successful to prevent the spreading of known attacks but leave a window of opportunity to attackers until the malicious website is added to the list and distributed to users [30, 36]. Therefore, other techniques were developed that include more proactive approaches. These are generally better at finding unknown phishing, but can have false positives and are not as widely used (e.g., integrated into browsers like Google Safe Browsing [18]). Here, Dou et al. compiled a list of approaches and their effectiveness [12]. Recently, machine-learning-based approaches to classify websites as phishing or benign based on features extracted from certificates have been proposed (e.g., [11, 24, 35]). Specifically, Dong et al. use and compare several machine learning algorithms to classify phishing websites [11]. They extract several features including information on the validity period and relation between subject and issuer fields. The best approach achieves a precision of more than 95%. Other machine learning approaches that focus on certificates, like the one by Torroledo et al., include features like the existence of several subject fields and validation levels [35]. Mensah et al. try to classify phishing and benign websites using features extracted from certificates and handshake information but conclude, that it is not possible to discriminate the two using only this type of information [24]. We come to a similar conclusion in that there are no general differences between certificates of benign and phishing websites. However, we go one step further by directly comparing the certificate of a phishing website to its target's certificate.

Even though these tools perform quite well (especially when compared to humans), there seems to be no solution employing these techniques widely, possibly due to the still rather high false positive rates. Unfortunately, the positive results do not translate well to user education: Not only do some features require complex computations to evaluate, but the classification process itself is also not applicable to users. In this paper, we extend the domain knowledge required to create effective classifiers by evaluating and arguing about certificate information as potential features.

Lastly, taking a look at **browser evaluation**, Biddle et al. set out to understand users' perceptions of the trustworthiness of a website when looking at certificates. They start with the assumption that users do not really understand certificates, and that the browser UI does therefore not help them make informed decisions. As such, they create an alternative UI for different validation levels and evaluate it in a user study. They find, that users' understanding of the original UI greatly varies, and that users do on average understand the level of trust a certificate provides better when using the proposed UI. Similarly, Sobey et al. also propose an alternative indicator for EV certificates [34]. They find using eye-tracking technology, that users did not notice the original EV indicators at all. However, the UI of Firefox has changed since then, making this information much more accessible (see Section 5). In this paper we analyze whether the certificate information relevant in the context of this paper is available to users in the browser

UIs and which steps users have to perform to get to this information.

## 4 Certificate Collection

This section describes the process and results of our certificate collection efforts in detail. The main goals are to answer research questions (1) and (2) as described in Section 1. We therefore look at general differences between the certificates of phishing and benign websites, as well as differences between the certificates of popular targets and their corresponding phishing websites. To achieve our goals, we collect certificate information from benign and phishing websites, extract features, and compare phishing and benign certificates.

### 4.1 Data Collection

For our analysis we retrieved 39 478 benign and 9 479 phishing certificates. In the following, we first describe how we collected benign and phishing domains and then describe, how we retrieved certificates from these domains.

#### 4.1.1 Data sources and preprocessing

In order to collect popular benign domains, we used the Alexa Top million list [1] and crawled the top 50 000 entries. Unfortunately, the Alexa data set does not include subdomains and for some domains, querying the domain without subdomains leads to a result that is different from querying the domain with its subdomains. A prominent example for this is PayPal, the most popular target for phishing campaigns in our data set. In this case, querying "`paypal.com`" leads to a certificate that differs from the one returned when querying "`www.paypal.com`". In order to mirror the experience of users more closely, we therefore apply a preprocessing step and query all benign websites using curl [7] to follow auto-redirects. We then use the resulting domain names for all further steps.

The phishing data set was obtained from Phishtank [29], a website that collects phishing websites collaboratively. Users can submit potential phishing websites and verify others, resulting in a peer-reviewed data set of phishing websites. However, this data set is not completely free of false positives: We did encounter some false positives when looking at specific certificates. We assume that this is due to one of the following reasons:

- The websites has been cleared and phishing content removed, but is still shown as "online and valid" by Phishtank.
- The websites were falsely flagged and the verification of users was wrong.

Either way, these cases seem to be rare in comparison to the data set of true phishing websites (we found less than ten cases in our detailed analysis in Section 4.2.2). We queried the Phishtank database for online and valid (i.e., verified by other users) phishing websites once daily over a period of 54 days (one day was missed due to technical problems). In this time, we collected 31 264 unique Phishtank entries.

#### 4.1.2 Certificate Collection

We use the following process to retrieve certificates from benign and phishing websites: First, we obtain the data sets for phishing and benign websites, using or converting to JSON representations of the data. For phishing websites, since we do not want to download certificates that have already been considered on a previous day, we merge the new data sets with a list of previously visited websites. Thus, we reduce the queried websites from several thousands to several hundred new phishing domains per day. This is not necessary for the larger benign data set, as these domains can be queried all at once.

After acquiring the websites to be queried, we start the crawling process using OpenSSL [28]. OpenSSL is an open source toolkit for the TLS and SSL protocols. We use the *s_client* component of OpenSSL to query websites and get certificate information [27]. The version of the program is "OpenSSL 1.1.1a FIPS 20 Nov 2018", as root certificates we utilize the Mozilla CA Certificate Store, which is, among others, also used by the Firefox browser [25]. We use *s_client* to connect to the specified domains on port 443 and retrieve a certificate, if possible. The certificates and additional information about the connection are saved on success for further analysis.

All in all, we were able to obtain 25 777 certificates from the 31 264 phishing domains. From these, we removed 11 712 duplicate certificates with respect to domain names in order to avoid polluting our data set with several entries for a single phishing campaign. To be precise, we create a database that only contains unique domain names and for each domain name exactly one certificate. This results in 14 065 certificates, but introduces a bias in our dataset, which now includes phishing campaigns using different subdomains, but disregards campaigns using different URL paths. Note that not all of the remaining certificates are unique. It is still possible that several different domain names are included in the same certificate. Next, we also decided only to look at certificates that are valid (as recognized by OpenSSL), since browsing to websites with invalid certificates generates a visible error in all major browsers to warn users. An overview of the validity status of phishing and benign certificates can be seen in Figure 1. *Name mismatch* errors (the domain name of the website does not match the subject `CN` or `SAN` of the certificate) were the most common, followed by *expired* certificates (validity period is in the past) and *self-signed* certificates. Overall,

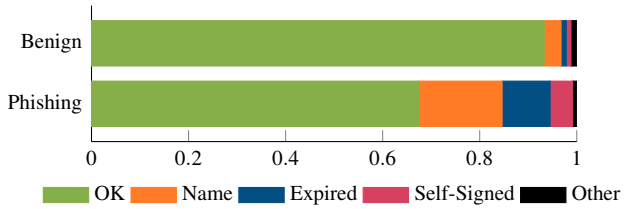



Figure 1: The validty status of certificates from phishing and benign websites.

| Field | |
|---|---|
| Subject: | `CN` |
| | `Organization` |
| CA: | Issuer `CN` |
| | Root `CN` |
| Validity: | Validity Period |
| | isValid |
| Extensions: | `SAN` |
| | Extended Validation |

Table 2: Features selected for further analysis.

phishing websites are more likely to present an invalid certificate than benign websites. Our final data set of valid phishing certificates contains 9 479 entries.

For benign websites, we remove 698 certificates with duplicate domain names and 2 842 invalid certificates and end up with a data set containing 39 478 benign certificates.

#### 4.1.3 Analysis and feature extraction

The analysis starts in a second pass, after all certificates are downloaded. Here, we scan all certificates, extract features of interest (see Table 2) and save them in a database. The features are divided into three groups:

- **Subject Information**: This group contains the subject `Organization` as well as validity and EV information. These are usually easily available to users and directly correspond to the websites a user might expect to be on.

- **Issuance Information**: This group contains the issuer and root `CN`, as well as the validity period. These are features that go beyond subject information, but are still easily available to users (see Section 5). We use the `CN` of the issuer rather than the `O` information of the issuer, as it is usually more detailed in our dataset.

- **URL information**: This group includes the subject `CN` and `SAN`, as well as the domain name of the website in question. We include this information to determine if looking at the certificate can be more effective than looking at the URL of a phishing website.

We disregard other fields commonly found in certificates for several reasons: Some fields are very similar (or the same) for all certificates issued by the same issuer (e.g., signature algorithm, policies). Thus, they only differ for certificates issued by different issuers, i.e. a field we already consider. Other fields consist only of long strings of numbers, that would be impractical to deal with in the context of user education and are unlikely to be usable in the context of automated phishing detection as well (e.g., public key, serial number). Lastly, some fields simply do not offer much variation at all (e.g., key usage, basic constraints).

### 4.2 Results

In the following, we first analyze the differences between our benign and phishing certificate collections w.r.t. to the information described in the last section and thus address research question 1. We then take a closer look at the differences between certificates of phishing websites and the certificates of their targets and thus continue with research question 2.

#### 4.2.1 General information in phishing and benign certificates

To address the first research question, we look at the distribution of features for benign and phishing certificates in general and try to find out how well they separate benign and phishing websites.

As described in Section 2, some CAs offer different levels of validation. It stands to reason that the more complex types of validation, i.e., organization and extended validation, make it harder for attackers to present a corresponding certificate. Still, we found that 1 444 certificates, about 15 % of all phishing certificates, include an `Organization` in their subject fields. We use the subject `O` field to decide if a certificate is OV, assuming that CAs follow best practices and do not include unverified information in the certificates they issue. For benign certificates, 13 852 or about 35 % of websites have an `Organization` in their subject fields. We assume that this difference is particularly pronounced for the higher ranks in the Alexa list, as these companies, with high user counts, have more incentives to buy organization validation or extended validation certificates. Taking this distribution into account, organization validation is not a deciding factor for differentiating phishing and benign websites, and would lead to many false positives if it were to be used as such.

EV certificates on the other hand are a more interesting matter. To decide if a certificate is EV, we use the subject `business category` field (OID: 2.5.4.15). This field is a requirement for EV certificates [5], checking for it is therefore an over approximation if we assume compliance to the CAB documents. We found that this approximation works quite well for otherwise valid certificates (we did not find any false classifications, even after working with and randomly

sampling our data set several times). Using this method, we identified only 39 phishing websites, that is about 0.4 %, with a valid EV certificate. These consist of compromised servers, as well as websites that are abused to host malicious content (e.g., `dropbox.com`, `jsfiddle.net`, `medium.com`), but also include several false positives (e.g., `paypal-notice.com`). We assume that such domains are less useful when phishing for user credentials, as they prominently display a different company in the URL bar of several popular browsers, but are still used for scams and other types of deception. As such, it seems that extended validation is less likely to be available to phishing websites, even though possible (social engineering) attacks were demonstrated before (e.g., [20]). Still, it is much harder to correctly fake an existing organization, including business registration details, as required for extended validation. On the other hand, only about 7 % (2746) of the benign websites use an extended validation certificate. Even among the top ten ranks, none protect their landing page with an extended validation certificate. This shows that even though an EV certificate (if it is valid and has the correct organization displayed) can be a good indicator that a website is legitimate, it does not provide a robust method to detect phishing websites. We found that some OV and EV certificates are used for phishing in connection with services that allow users to host content on their platforms. This includes Tumblr, Dropbox, Heroku and Medium. The interesting part of this phenomenon is, that these organizations have at least organization validated certificates. As such, a user that expects to be on bankingsite.com might open the certificate, look at the `Organization` and realize they are in fact on somehostingsite.com, which might awake suspicion.

The most popular issuers for benign and phishing websites are shown in Table 3. Again, we do not find any distinct features for phishing: the 10 most popular issuers, making up for 8598($\approx$ 90.7 %) of all phishing certificates, are also popular among benign websites (26046 certificates $\approx$ 66 %). As such, issuer information alone is not enough to separate benign from phishing domains. More detailed numbers for popular issuers for benign and phishing websites can be found in Tables 6 and 7 in Appendix A.

Similar to the issuers, we also find slight differences in other certificate details. The validity period for benign websites is on average longer than that of phishing websites (about 252 days for phishing and about 412 days for benign websites). We assume this is mainly due to the distribution of issuers: phishing websites more often use issuers with short validity periods like "Let's Encrypt" (90 days on average for both phishing and benign) and "cPanel"(average validity period of about 93 days for phishing, about 98 days for benign).

All in all, we do not find simple indicators for whether a certificate originates from a benign website or a phishing website. Attackers that set up their own websites have restrictions similar to benign administrators, resulting in similar choices for issuers and in similar certificates. Even though we found that phishing certificates often do not include an organization in the respective field, we found that this is also the case for many benign websites. The similarity in certificates is even more prominent if a benign website is used (compromised or not) to host an attacker's content.

| Issuer `CN` | Phishing | Benign |
|---|---|---|
| Let's Encrypt Authority X3 | 34.4 % | 17.4 % |
| cPanel, Inc. Certification Authority | 22.2 % | 1.6 % |
| RapidSSL TLS RSA CA G1 | 9.1 % | 0.2 % |
| COMODO RSA Domain Validation Secure Server CA | 5.3 % | 10.2 % |
| COMODO ECC Domain Validation Secure Server CA 2 | 5.2 % | 18.2 % |
| CloudFlare Inc ECC CA-2 | 5.0 % | 6.5 % |
| DigiCert SHA2 Secure Server CA | 3.4 % | 4.4 % |
| Go Daddy Secure Certificate Authority - G2 | 2.9 % | 4.4 % |
| Google Internet Authority G3 | 2.0 % | 0.5 % |
| RapidSSL RSA CA 2018 | 1.4 % | 2.6 % |

Table 3: Percentages of benign and phishing certificates issued by the 10 most popular issuers of phishing certificates.

### 4.2.2 Popular target websites

Next, we try to answer research question 2, i.e., the question whether the certificates of phishing websites differ significantly when comparing them to their target's certificate. For this, we look at the 15 most popular target websites of phishing attacks and their certificates (covering 2771 of 3275 valid phishing websites with a target label in the Phishtank database), and try to find out if and how well the phishing attacks are able to mimic their targets' certificates. We start by determining the login pages for the targets and noting their certificate information. Then, we look at the phishing data set and compare the target certificates with the phishing websites imitating these targets. The full results can be found in Table 4. Note, that all entries greater than one indicate unique domain names, that might still host several phishing websites on different URL paths.

First, we look at target organizations, and find that only few phishing websites are able to fake this information. To determine organization similarity we use the Python `difflib.SequenceMatcher` class [10], and manually verify all matches with a ratio of more than 0.3. We found no evidence of any phishing website obtaining a certificate with a spoofed organization name (even beyond the targets listed in Table 4). All entries in the table with a similar organization are hosted on the target's own infrastructure. For example, Microsoft offers several cloud services (e.g., Azure, SharePoint

| Target | Domain name | Number of phishing websites | Similar Organization | Same Issuer | Similar Issuer | Target in URL DN | Target matches wildcard |
|---|---|---|---|---|---|---|---|
| PayPal | www.paypal.com | 1169 | 0 | 1 | 24 | 84 | 12 |
| Facebook | www.facebook.com | 571 | 0 | 4 | 221 | 32 | 31 |
| Microsoft | login.live.com | 297 | 47* | 0 | 58 | 10 | 9 |
| ABSA Bank | www.absa.co.za | 214 | 0 | 0 | 0 | 5 | 0 |
| RuneScape | secure.runescape.com | 87 | 0 | 0 | 1 | 74 | 0 |
| eBay | signin.ebay.com | 67 | 0 | 1 | 0 | 5 | 0 |
| MyEtherWallet | www.myetherwallet.com | 62 | 0 | 1 | 2 | 15 | 0 |
| Blockchain | www.blockchain.com | 46 | 0 | 1 | 1 | 0 | 0 |
| Allegro | allegro.pl | 44 | 0 | 0 | 0 | 35 | 0 |
| Apple | appleid.apple.com | 42 | 0 | 0 | 2 | 8 | 3 |
| Steam | store.steampowered.com | 39 | 0 | 0 | 0 | 6 | 0 |
| Dropbox | www.dropbox.com | 37 | 1* | 1* | 0 | 2 | 1 |
| Binance | www.binance.com | 34 | 0 | 0 | 0 | 3 | 1 |
| Google | accounts.google.com | 33 | 1*[a] | 1* | 0 | 1 | 0 |
| ASB Bank Limited | online.asb.co.nz | 29 | 0 | 0 | 0 | 4 | 0 |

Table 4: Certificate and URL similarities for popular phishing targets. False Positives we found were removed. Entries marked with an asterisk are hosted on the target's own infrastructure.

[a] No text input, refers to different website

and OneDrive), that allow users to host content on domains owned by Microsoft. These domains are protected by Microsoft's own certificates and therefore match the target's `Organization`. We will encounter and argue more about this type of attack later on in this section.

Next, we look at issuers and their similarities to the target websites. The column "similar issuer" lists the number of phishing websites with similar issuers, meaning the same CA organization (e.g., DigiCert High Assurance is similar to DigiCert Extended Validation). We find, that many popular targets have few or no exact matches for the issuing CAs of phishing websites. Disregarding false positives and misclassifications again, only seven targets' issuers were replicated by phishing websites, and these cases are very rare (only one case for six targets, four for Facebook). Still, issuers seem to be a less precise metric than organizations as described above. This is also supported by the fact that there are many phishing websites with a similar issuer. It is also notable that among the 15 most popular targets we analyzed in detail, 9 are using EV certificates for their login pages. These require a thorough investigation of the entity requesting the certificate (see Section 2.3), making it less likely that organization information is spoofed. Looking at the details for similar and identical issuers reveals an interesting finding: Most of these entries come from websites that host user content, protecting it with their own certificate. In many such cases, users might still be able to recognize that they are not on the website they expect if they look at organization information. However, this is not the case if an attacker targets the service they are hosting their website on. We found this to be the case for Microsoft, as well as Google and Dropbox. To prevent such attacks, user content could be protected with a different certificate from the one used to login. Logins might be preferably protected with EV certificates.

Lastly, we look at URL similarities. We label a URL as similar to its target if it contains the organization or original domain name. We found that there are often far more similar URLs than either organizations or issuers. As shown in previous user studies (e.g., in [33]), complex phishing URLs can be difficult to detect even for users that were previously educated on the subject of phishing URLs. Interestingly, we find that attackers seem to be able to add the target name to the domain name in many cases (see Table 4). Therefore, even though many browsers offer a reduction in complexity by only showing the domain name, this part can still lead to users mistaking a phishing site for a benign site. As an aside, our database did not include a single valid certificate for a URL consisting of an IP-address, making this type of URL obfuscation less relevant than before (e.g., [16, 26]).

## 4.3 Discussion of collection results

We found that, unsurprisingly, there are no straightforward features extractable from certificates that instantly separate certificates of phishing and benign websites. We therefore answer research question 1 in the negative, concluding that there are no general differences between the certificates of phishing and benign websites. On the other hand, we found

that phishing websites currently do not seem to recreate the information included in the certificates of popular targets. So, as for research question 2, we find that currently there are some differences between the certificates of a phishing website and the certificate of its target. We particularly find that to date, the subject `O` and issuer `CN` seem not to be actively replicated.

On the other hand, it remains an open research question, whether it would be possible to expose the differences we observed to users in a way such that it would help them to detect phishing websites. In addition, it is unclear, how these differences could be used in the context of automated phishing detection.

Furthermore, while some information is currently not replicated, it is an open question how robust these findings are, i.e., how difficult it would be for an attacker to replicate the information on its target's certificate. Since organization validated certificates do not require the same level of vetting that EV certificates do, it is possible that attackers might get a fraudulent OV certificate without the risk of compromising their operations. It is also possible that a CA is compromised or misses a spoofed or fake organization in a certificate request. Replicating the issuer of a website is generally less complicated, as it does not require the attacker to spoof any information. As such, we conclude that it is not a robust feature to consider when analyzing a website.

In our analysis of popular target websites, we found that phishing websites with certificates that are similar to their target's certificate are often using hosting services and are not self-hosted. If the user content on such hosting services is protected by a wildcard certificate that includes information about the hosting service, it might still be possible to recognize this type of attack looking at the certificate. However, this is not the case if the service provider itself is the target.

Another potential problem with certificates as source of information for any future phishing detection tool is, that the tool might have problems with false positives if websites change their certificates. This includes a change from one issuing CA to another, or a change in validation level, both of which are possible scenarios for an organization.

A further potential problem of using certificates to detect phishing is, that automated tools (or users) might not be able to retrieve the certificate for a given website if they are affected by TLS interception [8, 14]. Here, the tool (or user) would only be able to retrieve the certificates of its interception middleware regardless of the website that is visited, rendering the detection of malicious websites with the help of certificate information entirely impossible.

For the sake of completeness, we also include some considerations that might have influenced our collection process. Firstly, our queries are performed from our country of residence, which might have influenced the results. This is more likely the case for larger websites that use content distribution and serve different content to users from different countries. Both phishing and benign websites were likely influenced by this, as attackers can use larger services to host their websites (see Section 4.2.2). This bias, however, is hard to remove and still represents a large amount of users that would have been served similar results.

Secondly, it is possible that attackers notice the crawling efforts and blacklist our client at some point (e.g., [26]). In this case, we would no longer be able to capture some of the attackers' methods, which might include more sophisticated techniques. We currently do not have any indications of this being the case.

In the next section, we will look at how certificate information is presented in popular browsers.

## 5 Browser Evaluation

In this section, we look at the presentation of certificates in major current browsers, taking into account the results from the previous section. There we looked at organization and issuer information included in certificates, as well as different levels of validation. In the following, we will compare the UI of several browsers with respect to their certificate presentation.

### 5.1 Browser UIs

We look at five browsers that cover a wide range of users: Google Chrome (Desktop[1] and Mobile[2]), Mozilla Firefox (Desktop[3]), Microsoft Edge (Desktop[4]), and Safari (Mac[5]).

First, we look at the browser's URL bars. We find that none of the browsers make a distinction between OV and DV certificates. However, all but Chrome for Android have special UIs for EV certificates. The indication for EV certificates ranges from additional information displayed next to the lock icon to more prominent highlighting of the lock and URL.

Next we count how many clicks are needed to get to subject `Organization` and issuer information. This varies greatly between the browsers. A comparison for non-EV certificates is given in Table 5. All browsers will open a smaller window after clicking on the lock (e.g., Figure 2), that includes general information about the current page.

The next part is where the browsers start to differ more prominently. All of the browsers offer a certificate viewer, that contains an overview of the certificate of the current website (e.g., Figure 3). We find that the subject `O` information is first available in the certificate viewer for all browsers we tested. Reaching this information takes different amounts of user input for the different browsers. While Edge and Chrome Desktop make this menu available after only one additional

[1] Chrome Desktop Version: 71.0.3578.98 (64-bit)
[2] Chrome Mobile Version: 74.0.3729.136
[3] Firefox Version: 64.0 (Build ID: 20181212110248)
[4] Edge Version: 42.17134.1.0
[5] Safari Version: 12.0.2 (13606.3.4.1.4)

| Browser | Subject O | Issuer CN or O |
|---|---|---|
| Chrome Desktop | 2 | 2 |
| Chrome Mobile | 3 | 2 |
| Edge | 2 | 2 |
| Firefox | 4 | 2 |
| Safari | 3 | 2 |

Table 5: Number of clicks required to get to subject and issuer information.

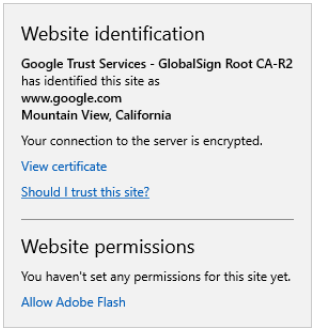



Figure 2: Pop-up after clicking on the lock symbol for an Organization validated website in Edge.

click, Firefox takes two more, for a total of four clicks, to open the certificate viewer. We also note that not all viewers are equally detailed, some are missing fields. For example, neither Edge nor Chrome Mobile includes information on extensions like certificate policies or basic constraints.

Note that the certificate viewers for Firefox, Chrome (Desktop and Mobile), and Safari offer an additional feature: The domain names are not translated from punycode, even if they are shown as IDN in the URL bar. This helps in preventing homograph attacks (e.g., [15]). We did not find websites that were translated to IDN in Edge's URL bar in the first place, making this less relevant in the case of Edge.

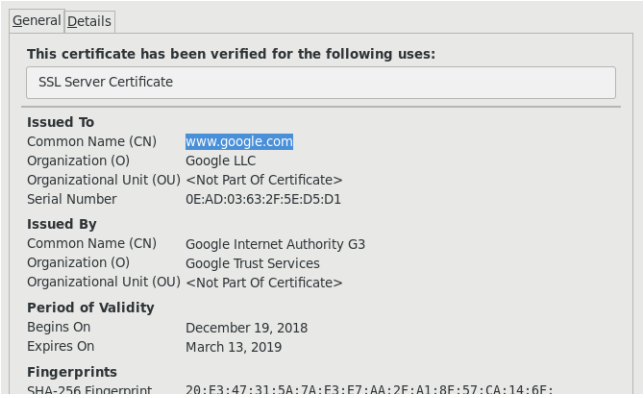



Figure 3: "General" tab of the certificate viewer in Firefox.

## 5.2 Discussion of Browser Evaluation

In Section 4.2.2 we looked at possible ways to recognize phishing websites, looking at issuer, subject and URL information. In this section, we discuss the certificate information presented by the different browsers in consideration of these findings.

We first make a distinction between EV and non-EV certificates, since the URL bar in most browsers is notably different for websites with EV certificates and those with non-EV certificates. In this case, some browsers (Edge, Chrome Desktop and Firefox) also show the subject `Organization` next to the URL, making the information readily available.

However, things are different for non-EV certificates. Here, no browser shows additional information by default without any user input. Only Edge displays some information after one click (issuer information and location if available), and all information discussed in this paper after two clicks. For Chrome Desktop, it takes users two clicks to get an overview of the certificate information, including `Organization` and issuer `CN`. Chrome Mobile requires an additional click to get to the certificate viewer, as does Safari. This is even more pronounced for Firefox: even though users will be able to verify the issuer `Organization` after two clicks, they will have to click through an additional window, four clicks in total, to get any information on the subject `Organization`.

Furthermore, some browsers did not include all fields of the certificate in their certificate viewer, though all of them contained the information discussed in this paper.

We also saw how hosting services can be abused and could offer a serious threat to unsuspecting users. Here, the browsers do include information about the current domain name, which might help mitigate the risk of hosting services.

All in all, we found that all of the fields discussed in this paper are available in all browsers we analyzed, yet this certificate information is available to users after different amounts of steps.

# 6 Conclusion and Future Work

Our analysis shows, that it is hard to differentiate phishing from benign websites using only information included in the certificate of a visited website, as certificates used by phishing websites include information that is very similar to that of benign websites, especially if both use certificates issued by the same issuer. This is plausible, considering the fact that phishers are often able to misuse the certificates of compromised servers, and that they will make decisions similar to the ones taken by administrators of benign websites when setting up their own servers.

We found that currently popular phishing targets often use EV certificates, and that it seems harder to copy websites using such certificates. Specifically, we found only a few instances of phishing websites where the issuer and organization of

the certificate used matched the equivalent information in the target's certificate. To assess if the differences we observed will persist in the future, we discussed how hard it would be for an attacker to obtain certificates that are more similar to their target's certificates. Unfortunately, it seems possible that at least some of the certificate features may be spoofed in the future.

Finally, we encountered instances of the particularly dangerous threat of hosting services, where user content is shown under the domain and protected by the certificate of a legitimate service. This can be abused by attackers to host their phishing websites, resulting in similar issuer and organization information as well as a similar URL on a legitimate looking top-level domain.

In future work, we plan to explore whether the observed differences between benign and phishing website certificates can be used to enhance the phishing detection capabilities of automated detection tools or users themselves. We also intend to further explore the question of how robust the subject `Organization` is against active attacks, and if subject spoofing might become more common in the future.

## Acknowledgments

This research was supported by the research training group "Human Centered Systems Security" sponsored by the state of North-Rhine Westphalia.

## References

[1] Alexa Top Sites. https://www.alexa.com/topsites. Online, accessed 26-Feb-2019.

[2] Mohamed Alsharnouby, Furkan Alaca, and Sonia Chiasson. Why phishing still works: User strategies for combating phishing attacks. *International Journal of Human-Computer Studies*, 82:69–82, 2015.

[3] Nalin Asanka Gamagedara Arachchilage, Steve Love, and Konstantin Beznosov. Phishing threat avoidance behaviour: An empirical investigation. *Computers in Human Behavior*, 60:185–197, 2016.

[4] CA-Browser Forum BR 1.6.3. https://cabforum.org/baseline-requirements-documents/, 2019.

[5] EV SSL Certificate Guidelines 1.6.8. https://cabforum.org/extended-validation/, 2018.

[6] Chromium Security: Marking HTTP As Non-Secure. https://www.chromium.org/Home/chromium-security/marking-http-as-non-secure. Online, accessed 27-Feb-2019.

[7] curl Website. https://curl.haxx.se/. Online, accessed 28-Feb-2019.

[8] X de Carné de Carnavalet and Mohammad Mannan. Killed by proxy: Analyzing client-end TLS interception software. In *Network and Distributed System Security Symposium*, 2016.

[9] Rachna Dhamija, J Doug Tygar, and Marti Hearst. Why phishing works. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 581–590. ACM, 2006.

[10] Python Docs: Difflib SequenceMatcher. https://docs.python.org/3/library/difflib.html. Online, accessed 24-Feb-2019.

[11] Zheng Dong, Apu Kapadia, Jim Blythe, and L Jean Camp. Beyond the lock icon: real-time detection of phishing websites using public key certificates. In *2015 APWG Symposium on Electronic Crime Research (eCrime)*, pages 1–12. IEEE, 2015.

[12] Zuochao Dou, Issa Khalil, Abdallah Khreishah, Ala Al-Fuqaha, and Mohsen Guizani. Systematization of Knowledge (SoK): A systematic review of software-based web phishing detection. *IEEE Communications Surveys & Tutorials*, 19(4):2797–2819, 2017.

[13] Julie S Downs, Mandy B Holbrook, and Lorrie Faith Cranor. Decision strategies and susceptibility to phishing. In *Proceedings of the second symposium on Usable privacy and security*, pages 79–90. ACM, 2006.

[14] Zakir Durumeric, Zane Ma, Drew Springall, Richard Barnes, Nick Sullivan, Elie Bursztein, Michael Bailey, J Alex Halderman, and Vern Paxson. The Security Impact of HTTPS Interception. In *Network and Distributed System Security Symposium*, 2017.

[15] Evgeniy Gabrilovich and Alex Gontmakher. The homograph attack. *Communications of the ACM*, 45(2):128, 2002.

[16] Sujata Garera, Niels Provos, Monica Chew, and Aviel D Rubin. A framework for detection and measurement of phishing attacks. In *Proceedings of the 2007 ACM workshop on Recurring malcode*, pages 1–8. ACM, 2007.

[17] Google Transparency Report: HTTPS encryption on the web. https://transparencyreport.google.com/https/overview. Online, accessed 28-Feb-2019.

[18] Google Safe Browsing. https://safebrowsing.google.com/. Online, accessed 22-Feb-2019.

[19] Anti-Phishing Working Group. Phishing Activity Trends Report: 3rd Quarter 2018. *APWG*, 2018.

[20] Collin Jackson, Daniel R Simon, Desney S Tan, and Adam Barth. An evaluation of extended validation and picture-in-picture phishing attacks. In *International Conference on Financial Cryptography and Data Security*, pages 281–293. Springer, 2007.

[21] J Klensin. Internationalized Domain Names in Applications (IDNA): Protocol. No. RFC 5891. Technical report, 2010.

[22] Ponnurangam Kumaraguru, Justin Cranshaw, Alessandro Acquisti, Lorrie Cranor, Jason Hong, Mary Ann Blair, and Theodore Pham. School of phish: a real-world evaluation of anti-phishing training. In *Proceedings of the 5th Symposium on Usable Privacy and Security*, page 3. ACM, 2009.

[23] Let's Encrypt. https://letsencrypt.org/. Online, accessed 26-Feb-2019.

[24] Pernelle Mensah, Gregory Blanc, Kazuya Okada, Daisuke Miyamoto, and Youki Kadobayashi. AJNA: Anti-phishing JS-based Visual Analysis, to Mitigate Users' Excessive Trust in SSL/TLS. In *2015 4th International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS)*, pages 74–84. IEEE, 2015.

[25] Mozilla CA Certificate Storage. https://www.mozilla.org/en-US/about/governance/policies/security-group/certs/. Online, accessed 24-Feb-2019.

[26] Adam Oest, Yeganeh Safei, Adam Doupé, Gail-Joon Ahn, Brad Wardman, and Gary Warner. Inside a phisher's mind: Understanding the anti-phishing ecosystem through phishing kit analysis. In *2018 APWG Symposium on Electronic Crime Research (eCrime)*, pages 1–12. IEEE, 2018.

[27] OpenSSL s_client manual page. https://www.openssl.org/docs/man1.1.1/man1/openssl-s_client.html. Online, accessed 24-Feb-2019.

[28] OpenSSL Official website. https://www.openssl.org/. Online, accessed 24-Feb-2019.

[29] Phishtank: Phishing Database. https://www.phishtank.com/. Online, accessed 26-Feb-2019.

[30] Swapan Purkait. Phishing counter measures and their effectiveness–literature review. *Information Management & Computer Security*, 20(5):382–420, 2012.

[31] Eric Rescorla. Http over tls, RFC 2818. Technical report, 2000.

[32] Steve Sheng, Mandy Holbrook, Ponnurangam Kumaraguru, Lorrie Faith Cranor, and Julie Downs. Who falls for phish?: a demographic analysis of phishing susceptibility and effectiveness of interventions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 373–382. ACM, 2010.

[33] Steve Sheng, Bryant Magnien, Ponnurangam Kumaraguru, Alessandro Acquisti, Lorrie Faith Cranor, Jason Hong, and Elizabeth Nunge. Anti-phishing phil: the design and evaluation of a game that teaches people not to fall for phish. In *Proceedings of the 3rd symposium on Usable privacy and security*, pages 88–99. ACM, 2007.

[34] Jennifer Sobey, Robert Biddle, Paul C Van Oorschot, and Andrew S Patrick. Exploring user reactions to new browser cues for extended validation certificates. In *European Symposium on Research in Computer Security*, pages 411–427. Springer, 2008.

[35] Ivan Torroledo, Luis David Camacho, and Alejandro Correa Bahnsen. Hunting Malicious TLS Certificates with Deep Neural Networks. In *Proceedings of the 11th ACM Workshop on Artificial Intelligence and Security*, pages 64–73. ACM, 2018.

[36] Yue Zhang, Serge Egelman, Lorrie Cranor, and Jason Hong. Phinding phish: Evaluating anti-phishing tools. *Network and Distributed System Security Symposium*, 2007.

## A Additional Results of Certificate Collection and Analysis

In the following we include several tables that contain additional details on our certificate collection results and its subsequent analysis. Tables 6 and 7 show the exact number of certificates issued by the most popular issuers for benign and phishing certificates.

| Issuer CN | Count |
|---|---|
| COMODO ECC Domain Validation Secure Server CA 2 | 7189 |
| Let's Encrypt Authority X3 | 6854 |
| COMODO RSA Domain Validation Secure Server CA | 4027 |
| CloudFlare Inc ECC CA-2 | 2564 |
| Amazon | 1908 |
| DigiCert SHA2 Secure Server CA | 1744 |
| Go Daddy Secure Certificate Authority - G2 | 1722 |
| GeoTrust RSA CA 2018 | 1426 |
| RapidSSL RSA CA 2018 | 1015 |
| DigiCert SHA2 Extended Validation Server CA | 1001 |
| GlobalSign Organization Validation CA - SHA256 - G2 | 825 |
| GlobalSign CloudSSL CA - SHA256 - G3 | 624 |
| cPanel, Inc. Certification Authority | 612 |
| DigiCert SHA2 High Assurance Server CA | 571 |
| COMODO RSA Organization Validation Secure Server CA | 523 |

Table 6: Number of benign certificates for the 15 most popular issuers.

| Issuer CN | Count |
|---|---|
| Let's Encrypt Authority X3 | 3259 |
| cPanel, Inc. Certification Authority | 2103 |
| RapidSSL TLS RSA CA G1 | 862 |
| COMODO RSA Domain Validation Secure Server CA | 502 |
| COMODO ECC Domain Validation Secure Server CA 2 | 489 |
| CloudFlare Inc ECC CA-2 | 474 |
| DigiCert SHA2 Secure Server CA | 321 |
| Go Daddy Secure Certificate Authority - G2 | 272 |
| Google Internet Authority G3 | 188 |
| RapidSSL RSA CA 2018 | 128 |
| Microsoft IT TLS CA 1 | 88 |
| GlobalSign CloudSSL CA - SHA256 - G3 | 74 |
| Actalis Domain Validation Server CA G1 | 70 |
| Amazon | 63 |
| DigiCert SHA2 High Assurance Server CA | 58 |

Table 7: Number of phishing certificates for the 15 most popular issuers.

# "We Can't Live Without Them!" App Developers' Adoption of Ad Networks and Their Considerations of Consumer Risks

Abraham H. Mhaidli, Yixin Zou, Florian Schaub
*University of Michigan School of Information*
*{mhaidli, yixinz, fschaub}@umich.edu*

## Abstract

Mobile ads pose privacy and security risks to consumers, including behavior tracking, malware, and inappropriate or biased content. Advertising networks connect mobile app developers with advertisers, enabling in-app advertising. We conducted a mixed-methods study with mobile app developers, consisting of survey and semi-structured interviews, to better understand why and how they partner with advertising networks, and their considerations of consumer risks in those interactions. Our findings focus on app developers who work independently or in smaller companies. We find that developers use advertising because they see it as the only viable way to monetize their app. Developers mostly choose an advertising network based on perceptions of which ad networks are popular rather than a holistic assessment. Despite claims of optimizing for profitability or consumer well-being, developers largely keep ad networks' default configurations. Developers are resigned to ad-related consumer risks, seeing themselves as unable to and not responsible for addressing the risks. Based on our findings, we discuss recommendations for mitigating consumer risks of mobile advertising.

## 1 Introduction

Many mobile apps use advertising to generate income [43]. Apps typically utilize *advertising networks* (e.g, Google AdMob, One by AOL, or Smaato [25]), which act as mediators between apps that are able to show ads and advertisers with ads to display. Ad networks provide revenue for apps; moreover, with ads apps can be offered free of charge, making them more broadly accessible.

However, ad networks are not without problems. In order to deliver relevant ads to users, ad networks use targeted advertising, for which they collect data about users through the apps or other means (e.g., online and app behavior, interests, geolocation, age, and gender) [65]. This pervasive data collection raises privacy concerns about access to this data, and whether it can be abused to manipulate or harm users [17]. Ad networks have been found to deliver offensive ads (which can emotionally harm users [65]) and discriminatory ads (e.g., promoting high paying jobs only to men or associating "black-identifying names" with prison sentences [20, 68]). Other ads have redirected users to malicious URLs that install malware onto users' devices [6, 51]. While issues with ads have been studied widely, we know little about how mobile app developers choose an ad network and to what extent they consider potential risks for their users in that decision.

Prior work on mobile developers' privacy and security behaviors found that developers want to choose 'good' third-party libraries, but may not be able to effectively evaluate them, e.g., because respective privacy policies are confusing [8]. Sources used to learn coding practices (e.g., Google vs. Stack Overflow) have also been shown to affect the security of resulting apps [1]. However, so far there has been no in-depth analysis of mobile ad network selection from the developers' perspective. Yet, understanding how app developers interact with and use ad networks is important for effectively tackling the consumer risks posed by ad networks. App developers have a crucial role in the in-app advertising ecosystem, as they decide whether and how they use in-app advertising.

To better understand app developer behaviors with ad networks, we investigated the following research questions: (1) Why do developers choose to monetize their apps through ads? (2) How and why do developers decide which ad network to use? (3) How do developers configure the ad networks they use? (4) How do developers manage the consumer risks posed by ad networks?

We conducted a mixed-methods study with mobile app developers, the majority of whom were independent app de-


*USENIX Symposium on Usable Privacy and Security (SOUPS)* 2019.
August 11–13, 2019, Santa Clara, CA, USA.

velopers. We surveyed 49 developers who have worked with in-app advertising and interviewed 10. We find that developers choose to use advertising out of a belief that it is the only viable way to monetize an app; most choose ad networks based on information in forums or a vague notion of which ad networks are used the most. Regarding their ad network configurations, developers claimed to optimize profit or ensure consumer well-being – however, we find that developers largely stuck to ad networks' default settings. While most developers were aware of consumer risks posed by ads, they were resigned about them. Most developers saw the responsibility to address those issues with ad networks, and viewed themselves as having little ability to effect change.

Our findings provide new insights on how app developers navigate the realm of in-app advertising. We conclude by discussing our findings' implications for intervention efforts to reduce the consumer risks of targeted ads and ad networks, including potential public policy directions and methods for better supporting developers in considering the implications of their ad network choices.

## 2 How Ad Networks Work

Advertising accounts for over half of mobile app revenue [43]. An advertising network connects *publishers* (i.e., app developers) and *advertisers* [50, 69]. Publishers offer ad networks space in their apps for advertising, e.g., a banner ad. The ad network pays the publisher a fee for this space, e.g., X dollars for every Y users who click an ad. The ad network then charges advertisers a slightly higher fee [50, 69]. The most commonly adopted ad network is Google AdMob [74], used by over 90% of apps that show ads [67]. Other popular ad networks include Facebook Audiences (9.86%), StartApp (8.82%), and Unity Ads (7.32%) [67].

### 2.1 Targeted Advertising

Ad networks often engage in *targeted advertising* [65], i.e., individual users are shown ads that are presumably relevant to their interests, e.g., someone who likes soccer might be shown an ad for tickets to a soccer match nearby [80]. The expectation is that since users are shown ads relevant to them, they are more likely to engage with the ad and buy the advertised product. Advertisers can select what groups of people should see their ads based on interest-profile selectors. This increases advertisers' revenue, and reduces resources spent on inefficient ads shown to consumers who are unlikely to engage with them [10]. Targeted advertising can benefit publishers as well: having ads that users are more likely to click increases the ad click rate and thus revenue [10]. Arguably, targeted advertising also benefits consumers, since consumers are not subjected to irrelevant ads [52].

However, targeted advertising also presents substantial risks for consumers. A necessity of ad targeting is the extensive gathering of information about individual consumers. This might include a consumer's online and app activities, age, gender, occupation, location, and other information inferred from individual behavior. This data is used to create a profile for a given individual. Ad networks infer what profiles are amenable to what sort of advertisements by monitoring who opens what kind of ads [52]. Information for ad targeting is often collected directly by ads displayed in an app and tracking code. For instance, when a user opens an app with an ad, the ad network code used to load the ad can (potentially) access the location information of the device, and so determine where the user is. The ad network can leverage this information to update the user's advertising profile and show the user more relevant ads (e.g., only showing ads for events near the user).

### 2.2 Ad Network Options for Publishers

To better understand how developers interact with ad networks, we analyzed the websites, terms of use, and documentation of five prominent ad networks [24]: Google AdMob, One by AOL, InMobi, Smaato, and StartApp. Overall we find that their services, functionalities, and even interfaces are very similar, with some minor differences.

To use an ad network to host ads in their app, developers apply for an account, and after review their account is approved or rejected. Once approved, developers have access to an online dashboard. Although dashboards differ among ad networks, they typically allow developers to view their revenue earned and the apps they have registered with the ad network. After registering an app, developers get access to the necessary code and IDs to integrate ads into their app. Integrating the ad code is fairly straightforward. Ad networks provide software development kits (SDKs) with which developers place ad display code in their app. The SDK typically allows developers to configure the type of ad to display (e.g., banner ad, video ad, etc.), its size, and where/when it appears in the app. Thus, while the ad network determines what ads get shown, the developer determines how ads are displayed.

Through the online dashboard, developers can further filter what ad categories may appear in their apps. Potential categories may include dating, cars, health, etc., and may differ by ad networks. By default, almost all categories are enabled. Google AdMob, though, has 'restricted ads' (for alcohol and gambling) that require publishers to opt-in to show them. Others (e.g., InMobi) have these same categories enabled by default. Developers can further block specific advertisers.

In addition, developers can choose (to some degree) what user data the ad network collects through a specific app by requesting certain mobile permissions for the app. Some permissions are required by the ad network (e.g., Internet connection, operating system, device type, network status); other permissions are not (e.g., precise geolocation). Developers can choose whether to provide this information to the ad net-

work. Other information developers can choose to send to the ad network include a user's age and gender, depending on what ad display code is used in the app. Developers have a financial incentive to share more information with ad networks: the more information is shared, the more relevant ads are delivered to users, and so, in theory, the developer's revenue will be greater. Lastly, targeted ads (as opposed to non-targeted ads) are the default option for all five ad networks studied, but developers can choose to display non-targeted ads.

# 3 Related Work

Prior research relevant to our work has focused on ad networks and developer behaviors regarding information seeking, tool selection, and privacy and security.

## 3.1 Consumer Risks of Ad Networks

Documented consumer risks posed by ad networks include (1) insensitive or offensive content [3]; (2) discriminating ads (e.g., ads for high paying jobs only shown to men, or ads that associate "black-identifying names" with criminal sentences and offer felony checks for individuals [20, 68]); (3) targeting based on sensitive content (e.g., religion) despite regulation against it [12]; and (4) excessive resource draining (such as battery and data) by ads [36, 57, 74]. Two prominent concerns are users' privacy and security.

Regarding users' privacy, ad networks collect information about users to target ads. This raises concerns over the vast quantities of information being collected, who has access to it, and for what purposes it is being used beyond advertising. Studies have found that ad networks collect extensive personal information, over-privilege apps to collect more information than needed, and that current protections are not effective at safeguarding user privacy [34, 46, 58]. A user's profile and data could be exposed, not only to an ad network and its advertisers, but to anyone who could access the ads seen by the user [14, 71]. Proposed solutions aim to protect consumer privacy while providing benefits of targeted ads [35, 38, 70], e.g., by performing targeting locally in the user's browser [70]. However, it is unclear how widely such solutions have been adopted.

Regarding users' security, a prominent risk is that of fraudulent ads that redirect users towards installing malware [27, 66], also known as madware [73]. Despite proposed solutions, such as improving malware classifiers using semantic features [15], it is still a prevalent problem. In 2017, Google AdMob purged over a billion ads, due to malware, phishing, and other consumer risks [63]. Additionally, there is the risk of sensitive ads being shown to users that can cause emotional discomfort or harm [65].

## 3.2 Developer Behaviors

To understand how app developers choose and engage with ad networks, it is important to know their information seeking behaviors. Social environment, especially information from colleagues and close friends, is highly influential in determining what tools developers adopt [39, 60, 79]. For instance, the adoption of security tools is heavily shaped by whether peers are utilizing that tool [56, 59, 78]. However, while peers are an important and useful resource for adopting new tools, they get used infrequently [55, 56].

Developers also use online forums and communities to find information and evaluate issues or complicated topics, such as code-related ethics, privacy risks, or the appropriateness of code contributions to a project [64, 72]. Trust and ease of access are major factors in determining what information sources developers use [39]. Contextual factors, such as familiarity with subject matter, stage of project, and client characteristics further impact information seeking behavior [28].

A common theme in studies of app developers' privacy and security behaviors is that developers often want to adhere to 'good' privacy and security practices (e.g., create secure code, respect user privacy), but fail to do so for a variety of reasons, such as lack of resources or expertise [5, 8, 32], faulty information sources [1], or insufficient documentation [26]. Balebako et al. found that app developers struggled to navigate complex privacy policies of third-party libraries, and were generally unaware of the data collected by such tools [8]. Egele et al. found that app developers often make mistakes when using cryptographic APIs [23]. Some app developers ask for more permissions than necessary, potentially for financial incentives [26, 46, 58]. However, Gorski et al. showed that API-integrated security advice can support developers in improving code security [32].

Despite the recognized privacy and security risks, very few studies have looked specifically at how developers choose ad networks. Some studies touch tangentially on this subject (e.g., Balebako et al. [8]) as part of more general investigations into developers' selection and use of tools. In contrast, our study provides deeper insights into both *how* and *why* app developers interact with ad networks, as well as to what extent and how they consider consumer risks in those interactions.

# 4 Study Design

To study mobile developers' behavior, practices, and attitudes regarding ad networks, we conducted a mixed-methods study involving a survey and semi-structured interviews. Our study was approved by the University of Michigan's IRB.

## 4.1 Survey

We first conducted an online survey to understand developers' attitudes and behaviors regarding ad networks (see Appendix

A). We asked about participants' experience developing apps and working with advertising networks. To gain more insights about particular experiences, we next asked participants to focus on one app for which they were involved in choosing and/or integrating ad network code. We asked what resources were used to choose an ad network, and had them rate which factors they valued when choosing an ad network. The survey concluded with demographic questions.

The survey was hosted on Qualtrics. Participants were given the option to enter a raffle for eight $20 Amazon gift cards. The median response time was 12.5 minutes.

### 4.2 Semi-structured Interviews

We conducted semi-structured interviews with some survey participants to gain deeper insights into mobile developers' views, behaviors, and attitudes towards ad networks (see Appendix B). We first asked participants about their background and experience with developing apps. Second, we asked which ad networks they had used and their respective experiences. Third, we asked how an ad network was chosen and how ads were configured in their app. Fourth, we asked about issues, problems, and consumer risks they had seen, heard of, or experienced with ad networks. Lastly, we asked broader questions to elicit their general thoughts regarding ad networks. Interview participants received a $15 Amazon Gift Card. Interviews lasted 27 to 42 minutes (median: 32 min.).

Interviews were transcribed and then analyzed with descriptive coding [53]. Two of the authors developed an initial codebook by jointly reading the transcripts and identifying emergent themes. They then iteratively refined the codebook by separately coding an interview, determining inter-rater reliability, revising the codebook as needed, and repeating with a separate interview. This procedure was repeated for 5 interviews until high inter-rater reliability was reached (Cohen's $\kappa$=.75) [53]. One of the authors then re-coded all interviews with the final codebook.

### 4.3 Recruitment

Our target population was app developers who have used ad networks in some capacity. Thus, our recruitment message asked for participants who had worked with ad networks, but did not mention privacy, security or risks.

We leveraged multiple channels to recruit participants, including posting in online forums aimed at app developers (e.g., the subreddit /r/AndroidDev) and technical Facebook groups; advertising through Craigslist; handing out flyers at local app developer meetups; reaching developers through personal contacts; and directly contacting developers based on contact information in the app store and LinkedIn.

The recruitment message advertised both the survey and interview component of the study, encouraging (but not requiring) participation in both. We conducted the survey and interviews in Summer and Fall 2018.

## 5 Findings

Our results show four key findings: (1) developers use advertising due to a belief that it is the only viable way to monetize an app; (2) when choosing an ad network, developers rely on online forums and ad networks' official websites, and do not spend much effort exploring which ad network to use; (3) developers often stick to ad networks' default configurations instead of optimizing for revenue or consumer safety; (4) developers do not view themselves as being able to or responsible for addressing consumer risks, believing that the responsibility lies with ad networks. We first discuss participant demographics, before presenting our findings in detail. We group findings by theme, combining findings from the survey and interviews, given that they address similar topics and complement each other nicely (quantitative information from the survey and rich qualitative insights from the interviews).

### 5.1 Participant Demographics

In total, 49 participants completed our survey. Their median age was 24 years (range: 18-47 years), which is relatively young compared to app developers' estimated average age (34 years) [45]. 37 participants were male, 3 female, 1 identified as non-binary, and the rest did not disclose their gender. This is reflective of the male-dominated app development field, e.g., over 90% of UK app developers are male [75].

Mobile app development experience varied: 8 participants had less than one year, 12 had 1-2 years, 9 had 2-3 years, 10 had 3-4 years, and 10 had more than 4 years of experience developing mobile apps. The median numbers of apps participants had worked on over the past three years was 6 (range: 1 to 100). Most (38) developed Android apps, 23 for iOS, and 13 for both; 1 developed for Windows Phone.

All participants provided the app's name and/or a link to its app store page. We analyzed each app's download and review numbers. As of May 2019, 26 apps were available in the Google Play store, with download numbers ranging from 10+ to 10,000,000+ (median: 10,000+), and review numbers ranging from 1 to >504k (median: 157). 12 apps were not available in the Google Play Store, but via APKPure, an alternative Android app market. 9 apps were in the iOS App Store, which only provides review numbers (range: 6-519k, median: 118). This snapshot shows that most participants' apps had a smaller audience, but other apps were highly popular.

About half of our participants (26) worked in small companies (four employees or less), and most (38) worked in small development teams (see Figure 1). This might be due to our recruitment strategies (e.g., directly contacting app developers via app store contact information), which were more likely to reach developers in small companies. However, it is reasonable to expect that developers in small companies/teams have

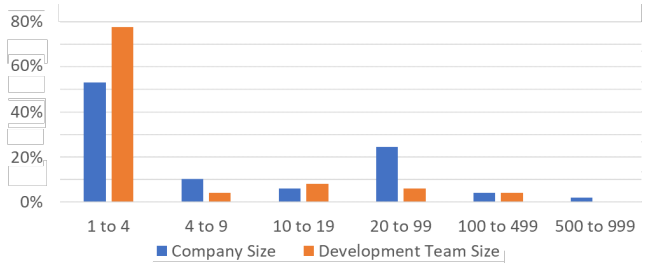



Figure 1: Company size (*blue*) and development team size (*orange*) of survey participants (n=49).

| | *Age* | *Years Experience* | *Ad Networks Used* | *Company Size (employees)* |
|---|---|---|---|---|
| P1 | 21 | 3–4 | Admob, StartApp, Flurry | 1–4 |
| P2 | 29 | > 5 | AdColony | 1–4 |
| P3 | 24 | 2–3 | Admob, StartApp, InMobi, Unity Ads | 10–19 |
| P4 | 22 | 1–2 | AdMob, ONE by AOL, Unity Ads | 1–4 |
| P5 | 39 | 1–2 | AdMob | 1–4 |
| P6 | 23 | > 5 | AdMob, StartApp, Unity Ads, Facebook Ads, Vungle | 20–99 |
| P7 | 19 | 3–4 | AdMob | 1–4 |
| P8 | 26 | < 1 | AdMob | 1–4 |
| P9 | 24 | > 5 | Admob, Facebook Ads, Vungle, App-O-Deal | 1–4 |
| P10 | 24 | 3–4 | AdMob, InMobi, Unity Ads | 1–4 |

Table 1: Interview participant demographics (n=10).

more say in how ad networks are chosen and used. Most (44) participants had worked as developers or software engineers; 23 as project managers; 16 as testers; 15 in upper management; 12 in marketing and 12 in user support (participants could select multiple roles).

9 survey participants were also interviewed. An additional interview participant (P9) did not fill out the survey, but contacted the researchers to participate in the interview directly. Table 1 provides their demographics. All interviewees were male; their median age was 24 years (range: 19–39 years). 8 participants were app developers working alone or in small teams (< 5 employees); P3 worked for a slightly larger company and P6 worked in upper management of a larger company that develops several apps.

## 5.2 Considerations in Adopting Advertising

Most participants used ads out of a resignation that ads are the only way to make money (despite general dissatisfaction with ad revenue), and after a careful evaluation of the type of app being developed. We first characterize ad network use before discussing why developers decided to use ads.

### 5.2.1 Ad Network Use Is Common

Ad networks were commonly used in the mobile apps developed by both survey and interview participants. 60% of the mobile apps developed by survey participants in the past three years used an ad network. 20 survey participants reported that advertising was the only monetization model they used. The most used ad network was Google AdMob (91% used it at least once), followed by Unity Ads (34%), inMobi (22%), and StartApp (20%). Others included Flurry (16%), Smaato (12%), One by AOL (12%), and LeadBolt (8%). This reflects a market dominated by Google AdMob, echoing prior work [74]. 16 survey participants and 6 interviewees had worked with three or more ad networks.

### 5.2.2 Resignation to ads as monetization model

When asked why advertising was chosen, 7 interview participants expressed a resignation towards advertising, saying it was the only viable way to monetize an app. They noted that because most apps are free (and monetized through ads), the only realistic way for an app to be competitive is to make it free as well. P10 said: *"I [knew] that many people wouldn't consider purchasing my app, so the only other viable option at the point seemed to include ads."* P9 was explicit: *"If it wasn't for advertising, almost all the independent developers would basically just die."* P6 mentioned a 'race to the bottom': when apps first came out they were expensive, but over time, app developers competed with one another, driving prices down and eventually forcing many apps to be free.

Despite a resignation towards ads, both survey and interview participants expressed dissatisfaction with ad revenue. In the survey, we asked participants whether and why they had changed ad networks in the past. All 10 survey participants who had changed ad networks indicated that higher revenue was a very/extremely important factor for changing – suggesting that their current revenue levels might be low or at least could be improved from their perspective. Moreover, 8 interview participants directly complained about the low revenue share they receive from ad networks. P7 stated: *"Google [AdMob] takes quite a big cuts of ad revenue obviously themselves, so you as the developer don't always see a lot of returns."* P2 similarly said: *"It's tough because the advertising dollars are so low, you need to have a large-scale viewership. You can't just have 1,000 people playing and watching."* While also disappointed with ad revenue, P10 had a different motivation for using ads: annoy users and encourage them to pay for the app's ad-free premium version.

### 5.2.3 Type of app matters for ad adoption

Alongside resignation with ads, all interview participants considered the type of app they were developing in their monetization choice. They would consider the app's genre, expected audience, and how often people would use the app and for

how long. Interviewees noted that for an app targeted toward a niche market, an app developer could charge users while still profit and gain traction. However, for a 'general' app with a wider audience, or an app that people would use infrequently, advertising was considered the only option for monetization.

#### 5.2.4 Showing ads to users considered fair

3 interviewees considered ads ultimately a fair way to monetize apps for both users and developers: for users, viewing ads may be annoying and inconvenient, but less so than paying for an app. P7, who used advertising in their app, said this is because the app offered a *"pretty basic service [which isn't] worth that much necessarily,"* and considered it unfair to charge users for it. Similarly, on comparing ads to charging for in-app purchases, P2 said *"I felt better about asking people to watch an ad rather than pay for a feature."* Another perspective emerging from the interviews is that ads were a fair compensation for the free app users were getting: if the app developer had spent significant time and energy creating an app, it was fair that users 'pay' the cost of seeing ads to compensate the developer.

Furthermore, 4 interviewees considered ads to have low impact on the user experience (less so than charging money). P1 justified his use of ads because *"[users] could always just shut off the ads and get rid of them,"* i.e., noting that it is up to the user whether they see ads or pay for the premium version.

### 5.3 Choosing an Ad Network

We asked both survey and interview participants how they selected the specific ad network for their app. In summary, participants either looked for information in online forums or acted on preconceived notions of what ad networks exist. This would lead them to a couple of ad networks, for which they would examine the ad network's website, and use it if it looked trustworthy. They typically kept using an ad network until it presented severe problems.

#### 5.3.1 Resources used to choose ad networks

Prior work suggests that developers often rely on friends and colleagues in tool selection [39, 60]. Our survey provided a different picture. Although 32% of survey participants rated friends as very/extremely important when choosing an ad network, the ad network's website (58%) and online discussion forums (45%) were rated as more important (see Figure 2).

The interviews revealed a more nuanced selection process. 8 interviewees reported choosing a particular ad network based on a vague awareness that other developers were using it with good experiences. P4 said: *"[What ad network to use] wasn't really a thing that we researched too heavily. It was more when we decided to kind of go that route, you're already kinda familiar with other people doing it; they seemed to have success with it, it didn't seem too difficult to add in."* For others, their ad network choice was based on what they had read in online forums, a vague awareness that a company existed, or even convenience (e.g., the ad network was supported by the SDK they were using to develop the app). 5 interviewees used rough heuristics to select an ad network. For example, 3 chose Google AdMob due to trust in its reliability, given that it is a large company. P5 chose AdMob because they believed it would work better on Android, given that Google develops Android. He said: *"Basically because it's Android, and as a Google product it seemed like the natural choice at the time because I trust them more. So I was like, 'Alright, I'll go with that.' And I've heard a lot about them so it made most sense."*

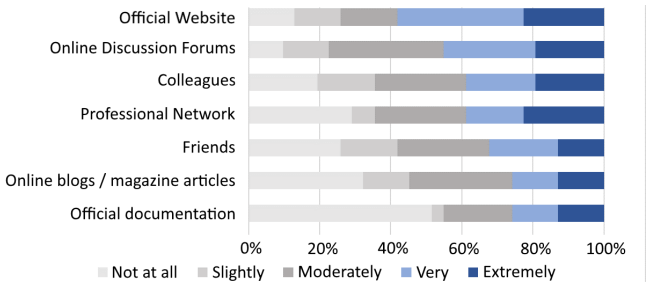



Figure 2: Survey responses to "How important were the following resources in deciding what advertising partner / advertising network to choose for APP NAME?"

6 interviewees reported they would then visit an ad network's website, and use the ad network if it looked trustworthy. Only 2 interviewees reported a conscious effort to compare and contrast different ad networks before choosing one.

#### 5.3.2 Sticking with a chosen ad network

Once they chose an ad network, most participants reported sticking with it. Only 20% (10) of survey participants had switched ad networks. "Competitor offering more revenue" was the most popular factor in this decision (3 rated 'very important;' 7 'extremely important'). Most interview participants (7) also stuck with their choice despite minor issues (e.g., low revenue), unless it posed severe problems or became unusable. Those who used ads in multiple apps typically used the same ad network for all apps, due to familiarity with the service and having all their revenue in one place.

#### 5.3.3 Exceptions for choosing ad networks

Among the 10 interviewees, P3 and P6, who both worked for larger companies with 20 or more employees, displayed unique patterns in ad networks selection. P3 was instructed by his company to use Google AdMob. Although he had no definite knowledge as to why Google AdMob was chosen, he

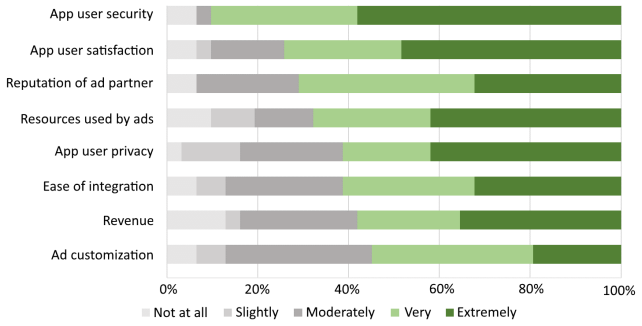



Figure 3: Survey answers to "In choosing an advertising partner / advertising network for APP NAME, what factors were considered, and how important were they in making the final decision?"

hypothesized that it was because past apps had used AdMob and the company had experience with it.

P6, as CEO of his company, would frequently switch ad networks so as to optimize revenue. In his words: *"What we would actually do is do an A/B test. [...] We would just write a particular logic method where when you first download an app [...] There would be two different [ad] networks integrated into the application, and randomly you would be assigned to one of them. We would get all the quantitative data. How often the ad shows, how often it fills, was it clicked? Was it annoying? Did the person delete it? What was the overall experience? What's the actual monetary vCPM [viewable cost-per-thousand impressions]?"* Yet, P6 noted that there were overall few changes in the set of 4–5 ad networks used, with AdMob being used the most.

#### 5.3.4 Factors considered in ad network selection

63% (31) of survey participants reported having been involved in ad network selection. Participants considered different factors in that decision (see Figure 3): 90% of participants considered the security of their users a very/extremely important factor. Similarly, 74% considered user satisfaction as very/extremely important. By contrast, revenue or ease of integration were valued less highly (58% and 61%, respectively). Least important were ad customization options.

However, revenue was the most popular decision factor for switching ad networks. 20% (10) of survey participants switched ad networks for an app, and for all revenue was a very/extremely important reason for switching. For half of them, revenue had been an equally important factor in their initial ad network choice. For the other half, revenue was more important in switching ad networks than initial selection.

The survey findings contrasted with the interview findings: 7 interviewees valued ease of integration the most, even if the ad network may have other shortcomings, such as revenue. P1 noted: *"[StartApp]'s definitely not the best, but it's just an ease to implement."* Considering that most interviewees were independent app developers, they might lack the resources to deal with complicated code, similar to previous findings that developers might lack the time to navigate complex privacy policies [8]. As P1 says *"I just don't have enough free time and I'd rather work on my app."* We note that for the interviewees who also took the survey, there were slight disparities between survey and interview responses in this respect. 2 participants who in the interview claimed that ease of integration was most important did not rate ease of integration as very/extremely important in the survey. Moreover, all who rated ease of integration as very/extremely important in the survey also rated at least one other factor (e.g., revenue) as extremely important in the survey, suggesting that ease of integration was on par with other factors.

The contrasting findings from the survey (where user satisfaction was valued highly in ad network choice, and revenue when switching) and the interviews (where ease of integration was valued highly in ad network choice) might be explained by social desirability bias. However, there were subtle indications in the interviews that the expressed care towards users is genuine, such as *"I don't want the app to be unfair to users"* (P7), or *"I felt better about not being intrusive to users"* (P5). Cognitive dissonance seems a more likely explanation: app developers want an ad network that does not harm their users, but integration and revenue take priority in practice, as they are factors directly experienced by the developers.

### 5.4 Sticking with Default Configurations

Despite claims of valuing certain factors over others, most interviewees (8) used ad networks' default ad settings and code options, *regardless* of the financial incentives. For instance, when using an ad network, developers can increase the amount of user data collected by the ad network by asking for additional mobile permissions, which in theory improves the relevance of ads shown to the user, and thus might enhance engagement and revenue. In contrast to past work finding that developers may add additional permissions for profit reasons [46], 9 interview participants claimed they used an ad network's default permissions or the bare minimum (only P9 added more permissions than necessary).

When asked if they used targeted or non-targeted advertising, 9 interviewees said they used targeted ads (the default), and 4 had not explored the possibility of non-targeted ads. The main reasons for using targeted ads were not only revenue increase (4), but also to provide a more enjoyable user experience (4), since users are not bothered by irrelevant ads. P7 said: *"I think [targeted ads are] more useful: for the developers, you end up making more money from them; and for the users seeing the ads, it's definitely more useful information."*

Moreover, most ad networks allow developers to customize what ad categories are shown in an app – by default all categories are enabled by most ad networks. 8 interviewees had

not changed the defaults; the other 2 restricted certain ad categories for apps aimed at children, or blocked a specific advertiser after a bad experience with them.

While developers explain configuration decisions with optimizing revenue or user experience, their configurations are often not consistent with the stated goals. For example, most interviewees rationalized targeted ads with improved revenue and user experience, yet they did not ask for additional permissions, which could further increase the accuracy of ad targeting (as well as increase privacy risks). Thus, rather than engaging in fully rational optimization, developers seem to be subject to status quo bias [62], even despite financial incentives to make adjustments.

#### 5.4.1 Projection onto app users as decision rationale

One interesting way developers rationalized their ad network configuration was to imagine themselves as the users. 4 interviewees would project themselves onto their users to decide what settings to use, using a logic of 'I don't like it when an app does X, so I will not do X to my users.' P5, in explaining why he chose to use banner ads, said *"I hate the ones that pop up and make you watch a video for thirty seconds because that breaks the flow of your app. I don't want them to interrupt, I just want to have extra content so banners made the most sense."* Similarly, P2 explained why he used a minimalistic banner ad in his app: *"If [users] don't want to, they can avoid it, and I think that's what is important to me personally as a player."* This again suggests that developers cared for and desired a good experience for their users. This care, though, is nuanced, in that it might have a financial aspect to it: an app that is harmful towards its users may lead to a decline in use. Thus, care is also important so as to maximize revenue.

### 5.5 Awareness of Consumer Risks

We asked survey participants to rate how true or false certain statements were, in order to assess their awareness of consumer risks posed by ad networks (see Figure 4). Overall, survey participants had mixed awareness of risks associated with ad networks. Ad networks have been found to sometimes collect user data without explicit consent [22]. When asked whether ad networks collect user data without users' permission, 41% of survey participants considered it probably/definitely true, but 20% false. Responses are skewed towards 'false' for ad networks' showing malicious ads (17 probably/definitely false; 5 probably/definitely true) or explicit and graphic ads (21 probably/definitely false; 10 probably/definitely true). Interviewees, on the other hand, were generally aware of ad networks' consumer risks, including malicious or graphic ads, privacy and data collection concerns, and excessive resource draining (battery and data).

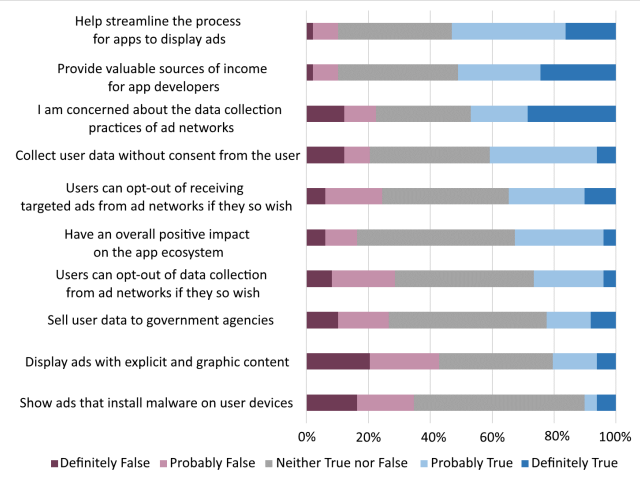



Figure 4: Survey participant answers to question "How much do you agree with the following statement: Advertising networks..."

#### 5.5.1 Awareness of risk does not lead to concern

Both survey and interview participants expressed mixed opinions about whether certain consumer risks were concerning. For instance, when asked if they were concerned about ad networks' data collection practices, 23 survey participants said definitely/probably true; 11 said definitely/probably false.

Interviewees' responses on what constitutes main consumer risks varied substantially. When asked what risks they were aware of broadly, 4 mentioned privacy concerns; P5 highlighted how ads used a lot of battery. We then probed interview participants about four specific types of risks: excessive data collection and tracking; graphic and inappropriate ads; malicious ads; and excessive resource draining. All interviewees claimed to have heard of these risks, but they had different opinions about how much they mattered. 6 interviewees described excessive resource draining as a minor risk, whereas the other four said it deserved attention, and one of them further described steps they had taken to mitigate this issue. Similarly, privacy was a big concern for some (4), but for the rest, such as P10, it was a minor issue: *"I don't view [privacy] as so much of a problem because it's not just the ad networks gathering it, it's almost all the major vendors and smartphones do that anyway to optimize their own services, so I think it's gonna happen either way."*

#### 5.5.2 Positive impression of ad networks despite issues

Despite acknowledging issues with ad networks, both survey and interview participants generally had a positive impression of them. 33% (16) of survey participants said it was probably/definitely true that ad networks had a positive impact on the mobile app ecosystem, as opposed to only 7 for proba-

bly/definitely false. Similarly, half (25) of survey participants said it was probably/definitely true that ad networks provide valuable sources of income for app developers. Interviewees were also generally favorable toward ad networks, commenting on how they help *"monetize an app that's not necessarily monetizable"* (P10), and on the ease and convenience they offer developers.

## 5.6 Managing Consumer Risks

Most survey and interview participants considered it the ad network's responsibility to manage and address consumer risks. They did not view themselves as having the agency to effect change in this regard.

### 5.6.1 Ad networks responsible for mitigating risks

Both survey and interview participants considered the ad networks responsible for managing and mitigating advertising-related consumer risks. Almost half (22) of survey participants said the ad network should be 'completely responsible' for removing bad ads found on ad networks; whereas only 2 survey participants considered app developers responsible, and 4 pointed at government regulatory entities. Similarly, all interviewees thought the ad network should be mostly, if not exclusively, responsible for addressing such problems, given the lack of control app developers have over what ads are shown in their apps.

Additionally, interviewees expressed resignation toward these risks and that app developers could do little to address them. When asked about issues with ad networks collecting excessive data, P10 talked about his inability to do anything about it: *"There is not much else I can do about it except not use the advertising service, but that's not really a solution. I don't even know what I could do to counter it."* Some interview participants further expressed an inherent trust in ad networks, in that these companies had the tools, willpower, and capability to filter out 'bad' ads, and thus there was little to worry on behalf of the app developers.

### 5.6.2 Little monitoring of ads

We asked interviewees whether they monitored the ads in their apps. P3 and P6's companies checked their apps frequently, by having dedicated employees or even a team systematically use the app to ensure the ads that appear are not malicious or explicit. However, most interviewees (8) did not make much effort to check if the ads in their apps were problematic. 3 explicitly stated that they did not monitor the ads in their app. For the 5 who did monitor the ads in their apps, they did so in a fairly informal way, such as using the app on a friend's device and seeing what ads appeared. This was similar to how P3 and P6 monitored their apps, but done in a much less frequent, structured and systematic manner.

3 interviewees explained this lack of monitoring with a fear of getting banned from an ad network. Most ad networks have measures in place to prevent automated or falsified clicks, i.e., 'clickfraud.' App developers found engaging in clickfraud may face penalties. Therefore, it is difficult for developers to check the ads in their apps without risking being reprimanded. P5 explained: *"I haven't personally [monitored my ads] because there are really strict rules, with AdMob, about triggering your own ads, because if you do that then it's kind of like trying to make your own money which is a problem."*

2 interviewees pointed out that it would be difficult to monitor malicious ads appearing in an app, given that ad selection is targeted to individual users. When talking about the possibility of viewers being exposed to overly graphic or explicit ads, P2 said *"If a person is targeted with ads that are more graphic in nature, the user would like it because [...] it's based on their viewing history."* This demonstrates strong trust in the accuracy of ad targeting and a disregard for the possibility of misuse or algorithmic bias.

# 6 Discussion

Our findings provide insights on (1) how developers choose ad networks, (2) how developers use and configure ad networks, and (3) how developers manage consumer risks posed by ad networks. When choosing an ad network, most participants feel resigned to the use of ads, viewing it as the only viable way to monetize an app. When configuring ad network settings, most participants used default settings *regardless* of the financial advantages or disadvantages of that choice. With respect to managing the consumer risks posed by ads, app developers are generally aware of the risks, but consider ad networks responsible for addressing them.

We first discuss limitations of our study, followed by opportunities for future research and intervention design that can better support developers in choosing and using ad networks, in ways that monetize their apps while mitigate consumer risks.

## 6.1 Limitations

Our study's sample consisted largely of developers working independently or for small companies (< 5 employees). This population constitutes an important faction of the app development ecosystem. For example, in France and Germany over 30% of app development companies had fewer than 5 employees, and in the UK, it is above 50% [76]. The differences in our interviews between small independent developers and one developer working for a larger company indicate that our findings are likely specific to small independent developers. Differences in ad network use between small and large app development companies should be studied in more detail in future research. Studying developers for small apps alone still provides a useful perspective though, given that most apps in

the market come from relatively new developers, and we need insight into the perspectives on ad-based monetization (and monetization more generally) from these individuals who just entered the field. Research with computer science students could also provide valuable insights in this regard.

Due to the specificity of our target population (app developers who had experience working with ad networks), our sample size is seemingly small (49 survey participants, 10 interviewees). However, our sample size is comparable to other studies examining software developer behaviors [9, 11, 16, 29, 30], due to general difficulties in recruiting participants who are professionals. Our study still provides rich insights into how small independent app developers manage ad networks and reason about associated consumer risks.

A common limitation in survey and interview studies relates to how participants may self-report behavior. Participants may not remember all details accurately, or may try to present a better self-image due to social desirability bias. We designed our survey and interview protocols in ways that avoid biasing participants. We also discussed potential indicators of social desirability bias in our findings.

## 6.2 Supporting Developers in Choosing App Monetization Models

Our findings suggest that many small app developers use ad networks out of resignation that advertising is the only way to make money from their app. Meanwhile, many participants complained that ad revenue was often low. It is questionable whether this resignation is well-founded: there are apps that exist without ads, and there seems to be little evidence to suggest that advertising is the only or most profitable way to monetize an app. Factors such as app category or what platform the app is on can influence how successful different monetization models are [7, 37, 61]. For instance, Roma et al. find that in Apple's App Store, paid and freemium monetization models generated higher revenues than free models, but they did not find significant differences between monetization models in the Google Play store [61]. Vratonjic et al. suggest that instead of adopting a blanket monetization approach, companies should strategically apply different funding approaches for individual users to maximize profits (e.g., using models to predict different users needs and wants, and serving ad-financed or fee-financed apps to different users) [77].

Given that our participants displayed limited knowledge of monetization models, we suggest a possible intervention: **presenting developers with more accurate information about what monetization models are available and optimal for an app under what circumstances, as well as associated risks or benefits for consumers**. This could increase developers' awareness of potential monetization models beyond the dominant reliance on advertising, and could encourage developers to adopt monetization models that increase revenue and pose fewer risks to consumers.

To accomplish this, more research is needed to (1) characterize and understand what monetization models are optimal for mobile apps under what circumstances; (2) analyze the impact of different monetization models on consumers, e.g., risks associated with each model and how consumers perceive them; and ambitiously, (3) explore new monetization models for apps that go beyond advertising and paid models, which ideally retain the low barrier to entry that free apps have, but do not pose the same consumer risks as targeted advertising.

One alternative way to finance apps is through crowdfunding, which has been an effective way to raise funds for projects related to games and journalism [4, 47]. This funding model could change the dynamic between app developer and consumer, creating a closer relationship and encouraging developers to act more responsibly towards their consumers [4, 13]. Another option for monetizing apps could lie in virtual currencies. For example, the social media platform Steemit rewards users who generate appreciated content with its own cryptocurrency: user accounts on Steemit are able to upvote posts and comments, and authors who get upvoted are rewarded with cryptocurrency tokens [18, 48]. Other platforms that adopt similar blockchain-based monetization models include Brave, SoMee.Social, Minds.com, and Presearch.org [2]. Moreover, certain subscription services (such as Youtube Premium) [31] work by having users pay a monthly flat fee, which gives users access to all content on their platform: the total money from these fees is distributed to the creators based on how much users interact with them (more interaction = larger share of the total money). Applying this to the context of mobile apps, one can imagine apps being monetized and valued based on the amount of downloads or users they have.

Once it is better understood which monetization models work best under what circumstances, as well as their respective benefits and disadvantages for both developers and consumers, a system (e.g. a website) could be constructed to aid developers with choosing a suitable monetization model for their app: after developers enter the characteristics of the app, such as the app's category and expected audience, the system would then recommend monetization models and show comparisons along multiple dimensions (e.g., revenue, user signup/conversion rate, public perception, and consumer risks). This system could be a standalone website or be offered by mobile platforms as part of their developer resources. It could also be integrated into online app development tutorials and courses (e.g., as a module on "financing your app"), as well as into integrated development environments (IDEs).

We argue that aiding developers with information grounded in research and data, as opposed to intuition or heuristics, could benefit both developers and consumers by highlighting less well-known monetization models with fewer consumer risks than advertising. The potential for success of this approach is supported by our finding that developers already engage in a deliberation process regarding their app's monetization model, but often in an unstructured manner. This

indicates that developers may be amenable to and benefit from more systematic information on monetization models.

### 6.3 Rethinking Ad Network Defaults

Participants in our study exhibited status quo bias [62]: they tended to stick to ad networks' default settings, regardless of the financial incentives involved. This implies that if harmful content appears in an app (e.g., sensitive products are being advertised), this is more likely due to the ad network's default setting, rather than any initiative by the developer. However, previous research has found that app developers sometimes ask for more permissions than necessary in their apps for financial reasons [46].

We thus propose that one way to limit consumer risks posed by ad networks could be **encouraging or mandating ad networks to change what the default settings are**. This approach has been used successfully in other contexts, such as healthy meal selection [41]. In the context of in-app advertising, the specific default settings to be regulated could relate to what permissions are set, whether targeted (or non-targeted) ads are used by default, and what categories of ads are permissible.

For instance, the default permissions required by an ad network could be reduced to the minimum necessary for the ad network to function. This would limit what data about consumers is collected and used for advertising purposes, and would also correspond to the GDPR's "data protection by default" principle. Additionally, in order to address privacy concerns of targeted advertising [80], the default could be set to 'non-targeted' rather than 'targeted' ads. Alternatively, it could be mandated that apps have to ask users for explicit consent to engage in targeted advertising (and if a user does not consent, show non-targeted ads). Consequently, fewer apps may engage in targeted advertising, perhaps alleviating some of the associated concerns. Moreover, currently it is common practice for most or all ad categories to be enabled by default. This should be changed so that certain sensitive ads, such as those for harmful products like tobacco or alcohol, political ads, or predatory ads (e.g., 'Get Rich Quick' ads that prey on vulnerable populations), are blocked by default. This could reduce the instances of such ads appearing and causing negative consequences, such as discomfort for consumers [3], the manipulation of people's voting behavior [42], and the sale of respective harmful products.

Of course, ad networks may be resistant to our proposed changes. There are financial incentives for maintaining the current defaults. Targeted ads may increase profit for the ad network, and greater data tracking may allow better (and so more profitable) targeted advertising [10]. Aside from the profitability of the ads themselves, more data might also hold better value for sale to third parties, such as data brokers. It is unlikely that ad networks will simply change their behavior due to these competing incentives, especially given that the advertising industry is mostly self-regulated through entities like the Digital Advertising Alliance.

We suggest regulators need to hold ad networks accountable by prescribing how defaults should be set up when self-regulatory approaches are ineffective. Consumer concerns about privacy risks are high [21], indicating that there may be political will to enact legislation. For instance, a recent report by the U.S. Government Accountability Office recommended that U.S. Congress should enact legislation to better protect consumers [19]. Other legislative efforts to regulate data tracking, such as the GDPR and the California Consumer Privacy Act (CCPA), have already been ratified and are being implemented. It is conceivable that future privacy legislation, such as the European ePrivacy Regulation or possibly a U.S. federal consumer privacy law, could stipulate more consumer-friendly default practices by ad networks.

App developers could also potentially drive ad networks to change defaults. App developers may desire to protect users, as directly suggested by our findings. Therefore, a collective call from app developers may exert pressure on ad networks. For instance, app developers could advocate that current app store requirements should be modified to avoid harmful content and prevent excessive data collection by default.

Finally, we should not neglect the possibility that ad networks may display goodwill. Faced with increasing concern and scrutiny surrounding data tracking practices, ad networks might want to regain consumer trust. Ad networks could set defaults that safeguard consumers to portray themselves as taking consumer safety and privacy seriously, while also providing a more explicit value proposition of targeted ads to consumers.

### 6.4 Encourage Developer Responsibility

Our findings indicate that developers care about the well-being of their users, e.g., most of our survey participants ranked app user security and satisfaction as very/extremely important in choosing an ad network. This aligns with Balebako et al.'s findings, suggesting developers want to create secure code that respects user privacy, but fail to do so for a variety of reasons such as struggling with complex privacy policies [8]. Our results reveal two main reasons why developers fail to mitigate ad-related consumer risks: (1) a belief that even though problems exist with ad networks, there is nothing app developers can do; and (2) a resignation that advertising is the only way to monetize an app.

Given this, we propose two opportunities for intervention. The first is to **correct the belief that developers cannot effect change**. At first glance, app developers may seem small when compared to ad networks, but they are still a crucial part of the advertising ecosystem. As such, they can effect change: both by simple actions such as configuring ads in certain ways (e.g., blocking ads for sensitive products), or more involved actions such as voicing complaints and concerns

over ad network practices, or boycotting certain ad networks. Second and more importantly, as a prerequisite of encouraging action, it is important to **make app developers realize that safeguarding app users from ad-related risks is not only the responsibility of ad networks, but also theirs**.

To encourage developers to take on responsibility, the focus of responsibility should be switched from blaming to collective action. Usually responsibility is talked about in terms of blame – if someone is responsible for consumer safety and the consumer is harmed, then that entity is blamed. Interpreting responsibility this way might be counter-productive, since it could alienate developers by painting them as 'guilty culprits.' Additionally, this interpretation does not show an accurate picture of the realities of in-app advertising. Loui and Miller discuss moral responsibility (as opposed to legal or causal responsibility) as a form of responsibility that, rather than seeking one actor or entity to blame for a system's problems, encourages all responsible actors to think critically about their role in the problem, and what they could do to mitigate the problem [49]. Similarly, Gotterbarn brings up 'positive responsibility,' a concept that does not seek to hold one party accountable or to blame for a system's problems, but rather motivates developers to think about the consequences of their actions on others [33].

Applying the positive responsibility framing to the context of in-app advertising, developers should not be blamed for the consumer risks of advertising. Rather, it emphasizes that in-app advertising is an ecosystem with multiple actors and stakeholders (advertisers, ad networks, app developers, and consumers). All members of the ecosystem do their part in allowing it to work, for good and for bad. The actions of those within the ecosystem influences how it will function – and so, it is on all the system's actors to make in-app advertising work better for everyone.

Given that developers seem to generally care about their users, as evidenced by our study and prior work [8], this suggests that developers might be amenable to taking actions that would mitigate consumer risks and protect their users. To achieve this, we suggest that it is important to show developers the power they have and the actions they can take. There are many places where this message could be promoted. One way is to target app development tutorials, courses, and online forums that developers visit frequently: creating new content that discusses positive responsibility and specific actions developers can take to mitigate consumer risks. This is in line with Mozilla's recent efforts to incorporate ethics into computer science curricula [54].

However, we acknowledge that encouraging developers to take responsibilities for consumer risks can be challenging. Not all developers would be willing to put in the effort needed to take on positive responsibility. Some might be in dire financial situations that make it difficult to properly care about their users. To address these barriers, material incentives should be created to encourage positive responsibility – perhaps a badge, token, or icon awarded to developers who proactively attempt to mitigate consumer risks of ad networks (e.g., a "fair trade" label for ads). Such a badge could be displayed to consumers as part of app descriptions, and help consumers identify responsibly designed apps. This would hopefully lead to more consumers using such apps, increasing their revenue, thus serving as an incentive for developers to earn this certification.

Even with these incentives, there are still challenges that positive responsibility faces. Further avenues of research could examine what factors could encourage the adoption of positive responsibility in developers, similar to research on encouraging other prosocial behavior (e.g., examining how economic incentives encourage blood donations or how technology can be used to increase empathy [40, 44]).

## 7 Conclusion

We conducted a mixed-methods study to better understand how and why developers choose and use ad networks, and how they manage consumer risks. We find that most developers feel resigned to use advertising, seeing it as the only viable way to profit from their apps. Developers mostly choose an ad network based on factors like which ad networks they perceive to be popular rather than a holistic assessment. Most developers use ad networks' default configurations *regardless* of the financial implications of that choice. Almost all developers believe the responsibility to mitigate the consumer risks of in-app advertising lies with ad networks.

We discuss several proposals for better supporting developers in mitigating consumer risks, such as presenting information on alternate monetization models for apps to developers, and enacting policy to make the default configurations of ad networks more consumer-friendly. Future work is needed to further explore these proposals, including both their effectiveness at overcoming consumer risks posed by in-app advertising, as well as challenges that we may face in getting developers to notice any provided guidance and support.

## Acknowledgments

This research has been partially funded by the University of Michigan School of Information. We thank Mark Ackerman, Allison McDonald and Steve Oney for feedback on early drafts of this work, as well as all members of the Security Privacy Interaction Lab (spilab) for their support.

## References

[1] Yasemin Acar, Michael Backes, Sascha Fahl, Doowon Kim, Michelle L. Mazurek, and Christian Stransky. You Get Where You're Looking for: The Impact of Informa-

tion Sources on Code Security. In *IEEE Symposium on Security and Privacy*, 2016.

[2] Activist Post. BRAVE: The Future Of Content Creation, Curation And PRIVATE Internet Browsing: Review. activistpost.com/2019/04/brave-the-future-of-content-creation-curation-and-private-internet-browsing.html, 2019. Retrieved 5/22/19.

[3] Lalit Agarwal, Nisheeth Shrivastava, Sharad Jaiswal, and Saurabh Panjwani. Do not embarrass: Re-examining user concerns for online tracking and advertising. In *Proc. of the 9th Symposium on Usable Privacy and Security*, pages 8:1–8:13, 2013.

[4] Tanja Aitamurto. The Impact of Crowdfunding on Journalism. *Journalism Practice*, 2011.

[5] Amirhossein Aleyasen, Oleksii Starov, Alyssa Phung Au, Allan Schiffman, and Jeff Shrager. On the privacy practices of just plain sites. In *Proc. of the 14th Workshop on Privacy in the Electronic Society*. ACM, 2015.

[6] Chaitrali Amrutkar, Kapil Singh, Arunabh Verma, and Patrick Traynor. Vulnerableme: Measuring systemic weaknesses in mobile browser security. In *International Conf. on Info. Systems Security*. Springer, 2012.

[7] Gil Appel, Barak Libai, Eitan Muller, and Ron Shachar. Retention and the Monetization of Apps. 2015.

[8] Rebecca Balebako, Abigail Marsh, Jialiu Lin, Jason Hong, and Lorrie Faith Cranor. The Privacy and Security Behaviors of Smartphone App Developers. In *Proc. of the Workshop on Usable Security*, 2014.

[9] Olga Baysal, Reid Holmes, and Michael W Godfrey. Developer dashboards: The need for qualitative analytics. *IEEE Software*, 30(4):46–52, 2013.

[10] Howard Beales. The Value of Behavioral Targeting. *Network Advertising Initiative*, 2010.

[11] Birgitta Bergvall-Kåreborn and Debra Howcroft. 'The future's bright, the future's mobile': A study of Apple and Google mobile application developers. *Work, Employment and Society*, 2013.

[12] Juan Miguel Carrascosa, Jakub Mikians, Ruben Cuevas, Vijay Erramilli, and Nikolaos Laoutaris. I always feel like somebody's watching me: Measuring online behavioural advertising. In *Proc. of the 11th Conf. on Emerging Networking Experiments and Technologies*, pages 13:1–13:13. ACM, 2015.

[13] Miguel Carvajal, José A. García-Avilés, and José L. González. Crowdfunding and Non-Profit media: The emergence of new models for public interest journalism. *Journalism Practice*, 2012.

[14] Claude Castelluccia, Mohamed-Ali Kaafar, and Minh-Dung Tran. Betrayed by your ads! In *Privacy Enhancing Technologies Symposium*. Springer, 2012.

[15] Wei Chen, David Aspinall, Andrew D. Gordon, Charles Sutton, and Igor Muttik. More semantics more robust: Improving android malware classifiers. In *Proc. of the 9th ACM Conf. on Security & Privacy in Wireless and Mobile Networks*, pages 147–158. ACM, 2016.

[16] Mauro Cherubini, Gina Venolia, Rob Deline, and Andrew J Ko. Let's Go to the Whiteboard: How and Why Software Developers Use Drawings. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*. ACM, 2007.

[17] Jeff Chester. Cookie wars: How new data profiling and targeting techniques threaten citizens and consumers in the "big data" era. In *European Data Protection: In Good Health?* 2012.

[18] Usman W Chohan. The concept and criticisms of steemit. *SSRN 3129410*, 2018.

[19] Catalin Cimpanu. Gao gives congress go-ahead for a gdpr-like privacy legislation. zdnet.com/article/gao-gives-congress-go-ahead-for-a-gdpr-like-privacy-legislation/, 2019. Retrieved 2/27/19.

[20] Amit Datta, Michael Carl Tschantz, and Anupam Datta. Automated experiments on ad privacy settings. *Proc. of Privacy Enhancing Technologies*, (1):92 – 112, 2015.

[21] Gary Davis. Key findings from our survey on identity theft, family safety and home network security. 2018.

[22] Soteris Demetriou, Whitney Merrill, Wei Yang, Aston Zhang, and Carl A. Gunter. Free for all! assessing user data exposure to advertising libraries on Android. In *Proc. of the Network and Distributed System Security Symposium*. Internet Society, 2016.

[23] Manuel Egele, David Brumley, Yanick Fratantonio, and Christopher Kruegel. An empirical study of cryptographic misuse in android applications. In *Proc. of the SIGSAC Conf. on Computer & Communications Security*, pages 73–84. ACM, 2013.

[24] TNS Experts. 12 popular mobile ad networks for app monetization. thenextscoop.com/mobile-ad-networks-app-monetization/. Retrieved 2/26/19.

[25] Alvaris Falcon. 20 Advertising Networks to Monetize Your Mobile App. hongkiat.com/blog/mobile-app-monetizing-networks/, 2017. Retrieved 07/24/2018.

[26] Adrienne Porter Felt, Erika Chin, Steve Hanna, Dawn Song, and David Wagner. Android permissions demystified. In *Proc. of the 18th ACM Conf. on Computer and Communications Security*, pages 627–638. ACM, 2011.

[27] Adrienne Porter Felt, Matthew Finifter, Erika Chin, Steve Hanna, and David Wagner. A survey of mobile malware in the wild. In *Proc. of the 1st ACM Workshop on Security and Privacy in Smartphones and Mobile Devices*, pages 3–14. ACM, 2011.

[28] Luanne Freund. Contextualizing the information-seeking behavior of software engineers. *Journal of the Association for Info. Science and Technology*, 66(8):1594–1605, 2014.

[29] Thomas Fritz and Gail C Murphy. Determining relevancy: how software developers determine relevant information in feeds. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, pages 1827–1830. ACM, 2011.

[30] Vahid Garousi and Tan Varma. A replicated survey of software testing practices in the Canadian province of Alberta: What has changed from 2004 to 2009? *Journal of Systems and Software*, 2010.

[31] Google. Your content & YouTube Premium. support.google.com/youtube/answer/6306276, 2019. Retrieved 06/05/2018.

[32] Peter Leo Gorski, Luigi Lo Iacono, Dominik Wermke, Christian Stransky, Sebastian Möller, Yasemin Acar, and Sascha Fahl. Developers deserve security warnings, too: On the effect of integrated security advice on cryptographic API misuse. In *Proc. of the 14th Symposium on Usable Privacy and Security*, pages 265–281. USENIX Association, 2018.

[33] Donald Gotterbarn. Informatics and Professional Responsibility. *Science and Engineering Ethics*, 2001.

[34] Michael C. Grace, Wu Zhou, Xuxian Jiang, and Ahmad-Reza Sadeghi. Unsafe exposure analysis of mobile in-app advertisements. In *Proc. of the 5th ACM Conf. on Security and Privacy in Wireless and Mobile Networks*, pages 101–112. ACM, 2012.

[35] Matthew Green, Watson Ladd, and Ian Miers. A Protocol for Privately Reporting Ad Impressions at Scale. In *Proc. of the SIGSAC Conf. on Computer and Communications Security*. ACM, 2016.

[36] Jiaping Gui, Stuart Mcilroy, Meiyappan Nagappan, and William GJ Halfond. Truth in advertising: The hidden cost of mobile ads for software developers. In *Proc. of the 37th International Conf. on Software Engineering*, volume 1, pages 100–110. IEEE, 2015.

[37] Daniel Halbheer, Florian Stahl, Oded Koenigsberg, and Donald R. Lehmann. Choosing a digital content strategy: How much should be free? *International Journal of Research in Marketing*, 2014.

[38] Michaela Hardt and Suman Nath. Privacy-aware personalization for mobile advertising. In *Proc. of the ACM Conf. on Computer and communications security*, 2012.

[39] Morten Hertzum. The importance of trust in software engineers' assessment and choice of information sources. *Information and Organization*, 12(1):1–18, 2002.

[40] Emma Nuraihan Mior Ibrahim and Chee Siang Ang. Communicating Empathy: Can Technology Intervention Promote Pro-Social Behavior?—Review and Perspectives. *Advanced Science Letters*, 2012.

[41] David R Just and Brian Wansink. Smarter Lunchrooms: Using Behavioral Economics to Improve Meal Selection. *Choices*, 24(3), 2009.

[42] Dan Keating, Kevin Schaul, and Leslie Shapiro. The facebook ads russians targeted at different groups, 2017.

[43] John Koetsier. 33% Of Mobile Revenue Now Delivered By Video Ads; Rewarded Video Is Most Effective. forbes.com/sites/johnkoetsier/2017/07/31/33-of-mobile-revenue-now-delivered-by-video-ads-rewarded-video-is-most-effective/, 2017. Retrieved 07/24/2018.

[44] Nicola Lacetera, Mario Macis, and Robert Slonim. Will there be blood? Incentives and displacement effects in pro-social behavior. *American Economic Journal: Economic Policy*, 2012.

[45] Simon Lee. 7 surprising statistics about the world of app development. thisisglance.com/7-surprising-statistics-about-the-world-of-app-development/. Retrieved 2/22/19.

[46] Ilias Leontiadis, Christos Efstratiou, Marco Picone, and Cecilia Mascolo. Don't kill my ads!: Balancing privacy in an ad-supported mobile application market. In *Proc. of the 12th Workshop on Mobile Computing Systems & Applications*, pages 2:1–2:6. ACM, 2012.

[47] Dario Lolli. 'The fate of Shenmue is in your hands now!': Kickstarter, video games and the financialization of crowdfunding. In *Convergence: The International Journal of Research into New Media Technologies*, 2018.

[48] Matthew Lopez. Steemit business model – how does steemit make money? feedough.com/steemit-business-model-how-does-steemit-make-money/, 2018. Retrieved 2/27/19.

[49] Michael C. Loui and Keith W. Miller. Ethics and Professional Responsibility in Computing. In *Wiley Encyclopedia of Computer Science and Engineering*. 2008.

[50] Ginny Marvin. What Is An Ad Network? martechtoday.com/martech-landscape-what-is-an-ad-network-157618, 2015. Retrieved 07/24/2018.

[51] Niels Provos Panayiotis Mavrommatis and Marf Monrose. All your iframes point to us. In *USENIX Security Symposium*. USENIX Association, 2008.

[52] Aleecia M Mcdonald and Lorrie Faith Cranor. Beliefs and Behaviors : Internet Users ' Understanding of Behavioral Advertising. *38th Research Conf. on Communication, Information and Internet Policy*, 2010.

[53] Mathew Miles, Micheal Huberman, and Johnny Saldana. *Qual. Data Analysis: A Methods Sourcebook*. 2014.

[54] Mozilla. Responsible computer science challenge. foundation.mozilla.org/en/initiatives/responsible-cs/, 2018. Retrieved 2/26/19.

[55] Emerson Murphy-Hill, Da Young Lee, Gail C. Murphy, and Joanna McGrenere. How do users discover new tools in software development and beyond? *Computer Supported Cooperative Work (CSCW)*, 24(5), Oct 2015.

[56] Emerson Murphy-Hill and Gail C. Murphy. Peer interaction effectively, yet infrequently, enables programmers to discover new tools. In *Proc. of the ACM Conf. on Computer Supported Cooperative Work*, pages 405–414. ACM, 2011.

[57] Abhinav Pathak, Y. Charlie Hu, and Ming Zhang. Where is the energy spent inside my app?: Fine grained energy accounting on smartphones with eprof. In *Proc. of the 7th ACM Euro. Conf. on Computer Systems*, pages 29–42. ACM, 2012.

[58] Paul Pearce, Adrienne Porter Felt, Gabriel Nunez, and David Wagner. Addroid: Privilege separation for applications and advertisers in android. In *Proc. of the 7th ACM Symposium on Information, Computer and Communications Security*, pages 71–72. ACM, 2012.

[59] Akond Rahman, Asif Partho, David Meder, and Laurie Williams. Which factors influence practitioners' usage of build automation tools? In *Proc. of the 3rd International Workshop on Rapid Continuous Software Engineering*, pages 20–26. IEEE Press, 2017.

[60] Mark A Robinson. An empirical analysis of engineers' information behaviors. *Journal of the American Society for Info. Science and Technology*, 61(4):640–658, 2010.

[61] Paolo Roma and Daniele Ragaglia. Revenue models, in-app purchase, and the app performance: Evidence from Apple's App Store and Google Play. *Electronic Commerce Research and Applications*, 2016.

[62] William Samuelson and Richard Zeckhauser. Status quo bias in decision making. *Journal of Risk and Uncertainty*, 1988.

[63] Stephen Shankland. Ads are great, Google says, except for the 3.2 billion bad ones. cnet.com/news/google-removes-billions-of-bad-ads-in-2017-bans-publishers/, 2018. Online; retrieved 09/03/2018.

[64] Katie Shilton and Daniel Greene. Linking platforms, practices, and developer ethics: Levers for privacy discourse in mobile application development. *Journal of Business Ethics*, Mar 2017.

[65] Sooel Son, Daehyeok Kim, and Vitaly Shmatikov. What Mobile Ads Know About Mobile Users. In *Proc. of the Network and Distributed System Security Symposium*. Internet Society, 2016.

[66] César Soto-Valero and Mabel González. Empirical study of malware diversity in major android markets. *Journal of Cyber Security Technology*, 2(2):51–74, 2018.

[67] Statista. Most popular installed mobile monetization software development kits (sdks) across global mobile apps in 2017. statista.com/statistics/742408/leading-mobile-app-monetization-sdks/. Retrieved 2/22/19.

[68] Latanya Sweeney. Discrimination in online ad delivery. *arXiv preprint arXiv:1301.6822*, 2013.

[69] Ian Thomas. Online Ad Business 101, Part III - Ad Networks. liesdamnedlies.com/2008/07/online-ad-bus-1.html, 2008. Retrieved 07/24/2018.

[70] Vincent Toubiana, Arvind Narayanan, Dan Boneh, Helen Nissenbaum, and Solon Barocas. Adnostic: Privacy Preserving Targeted Advertising. *Proc. of the Network and Distributed System Symposium*, 2010.

[71] Minh-Dung Tran. Privacy challenges in online targeted advertising. theses.fr/2014GRENM053/document, 2014. Retrieved 06/05/2019.

[72] Jason Tsay, Laura Dabbish, and James Herbsleb. Let's talk about it: Evaluating contributions through discussion in github. In *Proc. of the 22nd ACM SIGSOFT International Symposium on Foundations of Software Engineering*, pages 144–154. ACM, 2014.

[73] Bartlomiej Uscilowski. Mobile adware and malware analysis. symantec.com/content/en/us/enterprise/media/security_response/whitepapers/madware_and_malware_analysis.pdf, 2013. Online; retrieved 09/03/2018.

[74] Narseo Vallina-Rodriguez, Jay Shah, Alessandro Finamore, Yan Grunenberger, Konstantina Papagiannaki, Hamed Haddadi, and Jon Crowcroft. Breaking for commercials: Characterizing mobile advertising. In *Proc. of the Internet Measurement Conf.* ACM, 2012.

[75] Vision Mobile. Distribution of mobile app developers in the United Kingdom (UK) in 2014. statista.com/statistics/320488/app-developers-by-age-and-gender-uk/. Retrieved 2/22/19.

[76] Vision Mobile. Share of mobile app development companies with fewer than five employees in the United Kingdom (UK), Germany and France in 2014. statista.com/statistics/320476/app-companies-with-fewers-than-five-employees-uk-germany-france/. Retrieved 2/25/19.

[77] Nevena Vratonjic, Mohammad Hossein Manshaei, Jens Grossklags, and Jean Pierre Hubaux. Ad-Blocking games: Monetizing online content under the threat of ad avoidance. In *The Economics of Info. Security and Privacy*. 2013.

[78] Jim Witschey, Olga Zielinska, Allaire Welk, Emerson Murphy-Hill, Chris Mayhorn, and Thomas Zimmermann. Quantifying developers' adoption of security tools. In *Proc. of the 10th Joint Meeting on Foundations of Software Engineering*, pages 260–271. ACM, 2015.

[79] Shundan Xiao, Jim Witschey, and Emerson Murphy-Hill. Social influences on secure development tool adoption: Why security tools spread. In *Proc. of the 17th ACM Conf. on Computer Supported Cooperative Work*, pages 1095–1106. ACM, 2014.

[80] Frederik Zuiderveen Borgesius. Improving Privacy Protection in the Area of Behavioural Targeting. *SSRN*, 2015.

## A Survey Instruments

1. First, we would like to learn more about your experience as a mobile app developer. How many years have you worked in mobile app development?
   (a) Less than one year
   (b) Between 1 and 2 years
   (c) Between 2 and 3 years
   (d) Between 3 and 4 years
   (e) More than 5 years
2. What platforms have you developed apps for? (select all that apply):
   (a) Android
   (b) iOS
   (c) Blackberry
   (d) Windows Phone
   (e) Other (please specify)
3. How did you learn to develop mobile apps? (select all that apply):
   (a) Undergraduate major or course (e.g., BA in computer science)
   (b) Graduate major or course (e.g., masters degree in computer science)
   (c) Online course (e.g., a MOOC)
   (d) Self taught
   (e) Online tutorials
   (f) Workshop
   (g) On-the-job training
   (h) Other (please specify)
4. How many apps have you worked on in the last 3 years?
5. In the last 3 years, which role(s) have you carried out when working on mobile apps (select all that apply):
   (a) Developer, Programmer, or Software Engineer
   (b) Product or Project Manager
   (c) Tester or Quality Assurance
   (d) CEO or other high management / executive position
   (e) Sales / Marketing
   (f) User Support
   (g) Other (please specify)
6. Now we want to learn a little more about how you have integrated ads into apps. What role(s) have you played in regards to in-app advertising? Select all that apply.
   (a) I have been involved in choosing an advertising partner or advertising network for an app.
   (b) I have been involved in configuring the types of in-app ads shown in an app (e.g., where to place ads, what categories of ads to show, etc.)
   (c) I have been involved in integrating the necessary code into an app to enable in-app advertising
   (d) Other (please specify)
   (e) I have NEVER been involved in any way with regards to in-app advertising
7. Regarding mobile apps, have you used or worked with any advertising networks? if so, how often? (For each entry, participants answered to one of the following options: Have Not Used, Used in 1 app, Used in up to 3 apps, Used in up to 5 apps, Used in up to 10 apps, Used in more than 10 apps)
   (a) Google Ad Mob
   (b) ONE by Aol
   (c) InMobi
   (d) StartApp

(e) Smaato
(f) Flurry
(g) LeadBolt
(h) Unity Ads
(i) Other (please specify)

To learn more in-depth about your experience with in-app advertising, we want to ask you about a specific app you have worked in which you were especially involved with either choosing the advertising partner / advertising network to use, configuring what sort of ads are shown, or integrating the necessary code to display ads in the app.

8. Please name an app that utilizes in-app advertising and in which you were especially involved in decisions/integration regarding in-app advertising:
9. Please provide a link to this app in an app store/market (if unpublished enter N/A):
10. When did you work on the app?
11. What is the operating System for that app? Select all that apply.

    (a) Android
    (b) iOS
    (c) Blackberry
    (d) Windows Phone
    (e) Other (please specify)

12. Estimated company size for company that developed this app: (Options: 1-4, 10-19, 20-99, 100-499, 500-999, 1,000-4,999, 5,000-9,999, 10,000+)
13. Estimated development team size for team that developed this app: (Options: 1-4, 10-19, 20-99, 100-499, 500-999, 1,000+)
14. What role(s) did you have when developing this app? (select all that apply):

    (a) Developer, Programmer, or Software Engineer
    (b) Product or Project Manager
    (c) Tester or Quality Assurance
    (d) CEO or other high management / executive position
    (e) Sales / Marketing
    (f) User Support
    (g) Other (please specify)

15. How were you involved in the integration of ads into this app? (select all that apply):

    (a) I was involved in choosing the advertising partner(s) / advertising network(s) to use.
    (b) I was involved in deciding how ads are displayed in the app (e.g., where to place ads, what type of ads to show, etc.)
    (c) I was involved in integrating the ad network into the app
    (d) Other (please specify)

16. For each of the following role(s) with regards to in-app advertising, how involved were you in that role? (slider to the right = more involved)

    (a) Choosing what advertising partner / advertising network to use
    (b) Integrating the necessary code into an app to enable in-app advertising
    (c) Configuring the type of in-app ads shown (e.g., where to place ads, what categories of ads to show, etc.)

17. Revenue model of APP:

    (a) Free with In-App Advertising
    (b) Free with In-App Advertising, users can pay a fee to remove advertisements
    (c) Freemium model (app is free, certain features cost users money)
    (d) Paid download
    (e) In-App purchases (selling physical or virtual goods through the app)
    (f) Subscription (similar to Freemium, except instead of paying for extra features, users must pay for extra content)
    (g) Other (please specify):
    (h) Cannot remember

18. Who decided what revenue model to use in APP? (select all that apply):

    (a) Me
    (b) Programmer(s)
    (c) Project manager(s)
    (d) CEO and/or other upper level management
    (e) Investor(s)
    (f) Other (please specify):
    (g) I do not know who was involved in the decision process.

19. What ad formats does APP use? (select all that apply)

    (a) Banner ads (rectangular ads that occupy a portion of an app's layout; can be refreshed automatically after a period of time)
    (b) Interstitial ads (full-page ad format that appears at natural breaks and transitions, such as level completion in a game)
    (c) Native ads (advertisements presented to users via UI components that are native to the platform: for example, they can match the visual design of the app they are in)
    (d) Reward ads (Ad format that rewards users for watching ads)
    (e) Other (please specify)
    (f) Do not know / Cannot Remember

20. Who was responsible for choosing the ad formats used in APP? (select all that apply):

    (a) Me
    (b) Programmer(s) responsible for integrating the ad library code
    (c) Programmer(s) who were not responsible for integrating the ad library code
    (d) Project manager(s)
    (e) CEO and/or other upper level management
    (f) Investor(s)
    (g) Other (please specify):
    (h) I do not know who was involved in the decision process.

21. Which advertising networks, if any, were used in APP? (select all that apply):

    (a) Google Ad Mob

(b) ONE by Aol
(c) InMobi
(d) StartApp
(e) Smaato
(f) Flurry
(g) LeadBolt
(h) Unity Ads
(i) Other (please specify)
(j) No advertising network was used in this app.
(k) Cannot remember

22. Who decided what advertising partner / advertising network to use in APP? (select all that apply):

(a) Me
(b) Programmer(s)
(c) Project manager(s)
(d) CEO and/or other upper level management
(e) Investor(s)
(f) Other (please specify):
(g) I do not know who was involved in the decision process.

[If participant indicated they were involved in choosing an advertising network]

Your previous answers indicate that you were involved in selecting an advertising partner / advertising network for APP. These next questions will ask more about that process.

23. How important were the following resources in deciding what advertising partner / advertising network to choose for APP? [Not at all important, slightly important, moderately important, very important, extremely important, and an additional N/A option]

(a) Friends
(b) Colleagues (fellow developers/others internal to the company)
(c) Professional Network (fellow developers/others external to the company)
(d) Official website(s) of advertising partner / advertising network
(e) Official documentation and / or documents from advertising partners / advertising networks (e.g., SDK documentation, privacy policy, Terms of Service)
(f) Online blogs / magazine articles
(g) Online discussion forums (e.g., Reddit, StackOverflow)
(h) Other (please specify)

24. In choosing an advertising partner / advertising network for APP, what factors were considered, and how important were they in making the final decision? [Not at all important, slightly important, moderately important, very important, extremely important, and an additional N/A option]

(a) Revenue provided (e.g., eCPM rate)
(b) Ease of integration
(c) App user privacy
(d) Reputation of advertising partner / network
(e) Ad customization options offered (e.g., customize ad format, ad content, types of ads shown...)
(f) App user's security (e.g., likelihood of ads serving malware)
(g) App user's satisfaction / experience
(h) Resources used by ads (e.g., battery, network data)
(i) Other (please specify)

[Shown to all participants]

25. In what ways, if any, have the ads shown in APP been configured or customized? (select all that apply)

(a) Blocked certain advertisers / URLs
(b) Blocked certain categories of ads from being shown in the app
(c) Use only non-personalized or non-targeted ads
(d) Other (please specify)
(e) The ad content of APP has not been customized in any way
(f) I do not know if any configurations were made
(g) Prefer not to say

26. If the ads shown in APP were customized, who decided what configuration to use? (select all that apply):

(a) N/A / Ads were not customized
(b) Me
(c) Programmer(s) responsible for integrating the ad library code
(d) Programmer(s) who were not responsible for integrating the ad library code
(e) Project manager(s)
(f) CEO and/or other upper level management
(g) Investor(s)
(h) Other (please specify):
(i) I do not know who was involved in the decision process.

27. If decisions were made to configure the ad content, please explain why the ads in APP were configured this way? If the answer is not known, or not applicable, please respond N/A.
28. Some advertising partners / advertising networks collect data through the advertisements inside an app. What information does the advertising partner / advertising network used in APP collect or have access to? [Does have access, Probably has access, Probably does not have access, Does not have access, unsure]

(a) Device ID
(b) Operating System Information (e.g., what OS is on the device)
(c) Coarse Location
(d) Precise Location
(e) Age of user
(f) Gender of user
(g) Name of user
(h) Contact list of users
(i) Microphone
(j) Camera

29. Some advertising partners / advertising networks allow developers to customize what data is collected through the in-app ads and sent to an advertising partner / advertising network. If such customizations were made, who was in charge of that decision? Please select all that apply.

(a) Me

(b) Programmer(s) responsible for integrating the ad library code
(c) Programmer(s) who were not responsible for integrating the ad library code
(d) Project manager
(e) CEO and/or other upper level management.
(f) Investors
(g) Other (please specify):
(h) I do not know who was involved in the decision process.
(i) N/A / No customizations were made

30. Has APP experienced any of the following issues with regards to its advertising partners / advertising networks? If so, how often? [Never, Rarely, Occassionally, A moderate amount, A great deal, and an additional Unsure option]

(a) Failure to receive payment (or received late payment) from advertising partner / advertising network
(b) Advertising network account being deleted or banned without explanation.
(c) Inappropriate or undesired ads shown in app (e.g., an advertisement that displays explicit pornographic material, graphic violence)
(d) Malicious and/or harmful ads shown in app (e.g., advertisements that install malware onto user devices)
(e) Complaints from users about the type of ads shown in your app
(f) Advertisements not displaying in app
(g) Advertising network being slow or inefficient responding or addressing issues
(h) Being misled, lied to, or otherwise deceived by advertising network's policies and guidelines.
(i) Excessive data collection by advertising network
(j) Other (please specify)

31. If any of the above issues were experienced, please briefly describe what steps, if any, were taken to address them (if no steps were taken, please write N/A):
32. Did the advertising partner for APP change in the time you worked on this app?

(a) Yes
(b) No
(c) Do not know / Unsure

[Shown only if answer to Q32 was yes]
33. What reasons prompted the change of advertising partners / networks for APP, and how important were they in making the decision to change? [Not at all important, slightly important, moderately important, very important, extremely important, and an additional N/A option]

(a) Competitor offered better revenue (e.g., higher eCPM (effective cost per one thousand impressions) rates)
(b) Competitor offered more customization options (e.g., more customizability with regards to what ads to place)
(c) Competitor offered an overall better product (e.g., offered higher quality ads)
(d) Advertising partner / advertising network displayed ads that were harmful to users of the app (e.g., the ads installed malware on user devices).
(e) Advertising partner / advertising network displayed ads that were explicit (e.g., advertisements that showed pornography, ads that showed graphic violence).
(f) Other (Please Specify)

Now we want to learn a bit more about your perception of advertising networks.
34. For each of the following statements, please indicate to what degree you think the statement is true. [Definitely false, Probably false, Neither true nor false, Probably true, Definitely true]

(a) Advertising networks provide valuable sources of income for app developers.
(b) Advertising networks collect user data without consent from the user
(c) Users can opt-out of data collection from advertising networks if they so wish.
(d) Users can opt-out of receiving targeted advertisements from advertising networks if they so wish.
(e) Advertising networks help streamline the process for apps to display ads.
(f) Advertising networks show advertisements that install malware on user devices.
(g) Advertising networks show advertisements that show explicit and graphic advertisements (e.g., pornographic material, explicit violence and gore, etc.)
(h) Advertising networks sell user data to government agencies (e.g., the FBI)
(i) Advertising networks have an overall positive impact on the app ecosystem
(j) I am concerned about the data collection practices of advertising networks

Sometimes, when an app uses an advertising network to display ads, an ad can be shown that is either harmful to users (e.g., installs malware on user devices) or otherwise illegal (e.g., displays explicit ads for terrorism or prostitution).
35. If such a thing happens, who do you think SHOULD BE responsible for fixing the issue? [Not at all responsible, Somewhat responsible, Mostly responsible, and Completely responsible]

(a) Network
(b) App
(c) Government Agency (E.g., the FTC or the FBI)

36. If such a thing happens, who is currently responsible for fixing the issue? [Not at all responsible, Somewhat responsible, Mostly responsible, and Completely responsible]

(a) Network
(b) App
(c) Government Agency (E.g., the FTC or the FBI)

Thank you for your time! We are almost done. We would like to ask you to complete some basic demographic questions:
37. Please enter current age in years, in a number format (if you'd prefer not to say, enter 0):
38. What is your gender?

(a) Male
(b) Female
(c) Non-binary

(d) Other (please specify)
(e) Prefer not to say

39. Highest level of education achieved (if currently enrolled, highest degree received.):

(a) No schooling completed
(b) Some high school, no diploma
(c) High school graduate, diploma or the equivalent (for example: GED)
(d) Some college credit, no degree
(e) Trade/technical/vocational training
(f) Associate degree
(g) Bachelor's degree
(h) Master's degree
(i) Professional degree (e.g., J.D., M.D.)
(j) Doctorate degree
(k) Prefer Not To Say

40. Current employment status:

(a) Full time employment for salary / wages
(b) Part time employment for salary / wages
(c) Self-employed
(d) Unemployed
(e) A homemaker
(f) A student
(g) Retired
(h) Unable to work
(i) Prefer Not To Say

## B Interview Protocol

1. First, I'd like to learn more about your experience developing mobile apps. How did you get into mobile app development?
2. What is your current role? What role(s) have you played in the past? How has working on them been like?
3. Now I want to ask a bit more about in-app advertising. Can you describe the purpose of an ad network and how it functions to provide ads in your app?
4. What is your experience with advertising networks? Which ones have you worked with? What was this like? How long have you worked with ad networks? What role(s) have you played in working with them?
5. Now I want to talk in depth about a specific app you have worked on in which you were heavily involved in incorporating ads into your app. Describe a typical day working on APP. What was your role? How were ads used in APP? Was an ad network used? Do you remember which one? Why was it decided to use advertising in APP?

   I want to focus on your experiences with APP. But feel free to mention experiences you have had with other apps.
6. What were the reasons why advertising was used in APP? What is the business model of APP? Can you walk me through how the model was chosen? Was an alternative without advertisements considered? Why / Why not?
7. Walk me through the decision to use [AD NETWORK]. Who was involved in making that decision? What were you looking for in an ad network? I.e, what were your priorities? Why was this particular ad network chosen? How much time was spent researching each company? Were other ad networks considered? Why was [AD NETWORK] chosen over other ad networks? Are you using other ad networks in parallel? Do you utilize mediation?
8. What resources, if any, did you use to help choose what ad network to use? Walk me through how they were used.
9. Walk me through the process of integrating the ad library code into your app. How easy or difficult was it to integrate the ad library code into your app?
10. Some ad networks allow you to customize what permissions are needed for the app to function. Do you remember what permissions were set in APP? Why were they set / not? How did you (or your team) come to this decision?
11. Some ad networks collect data about the phone that uses the app. Do you know what data the ad network collects through APP?
12. With [AD NETWORK] you can choose what data is sent to the advertiser to deliver better targeted ads. Do you know what data is sent to the ad network through APP? Why / Why not? Do you see any issues with this sharing of data?
13. Do you know if the ads shown by your ad network are targeted? Do you know if you can change this to non-targeted ads? Have you ever explored the option of non-targeted ads? Which ones do you use? Why? What are your own views on targeted advertisements?
14. With [AD NETWORK] you can configure what category of ad your app shows – for example, you can choose where apps that show clothing appear in your app, or block a certain vendor or advertiser from your app. Have you ever blocked a certain category from your app? Why or why not? Have you ever blocked a specific advertiser from your app? Why or why not?
15. Have you had any experiences with ad networks that are different from the ones you have just described?
16. What are the main benefits you see with advertisement networks?
17. What are the main issues you see with ad networks?

    Some ad networks have been known to show advertisements that are offensive or harmful to users (e.g., ads that display pornographic or offensive material, ads that download malware onto user devices. . . )
18. Have you ever received complaints that there have been these bad ads on one of your apps? If so, how did you deal with them?
19. Have you ever checked the ads on your ad network for these issues? If so, how?
20. Do you know your ad networks policy on these sorts of ads?
21. If a harmful ad like this is found in an app, whose responsibility is it to remove it? Why?
22. Now I will walk through series of issues that have been identified with ad networks. I want to know if your company/team were aware of these issues, and if so, if any steps were taken to mitigate them? (Issues mentioned: malware, inappropriate / offensive content, battery draining due to sharing of data, using up user's mobile data plan, companies obtaining user data without explicit permission.)

# Usability Smells: An Analysis of Developers' Struggle With Crypto Libraries

Nikhil Patnaik
*University of Bristol*
*nikhil.patnaik@bristol.ac.uk*

Joseph Hallett
*University of Bristol*
*joseph.hallett@bristol.ac.uk*

Awais Rashid
*University of Bristol*
*awais.rashid@bristol.ac.uk*

## Abstract

Green and Smith propose ten principles to make cryptography libraries more usable [14], but to what extent do the libraries implement these principles? We undertook a thematic analysis of over 2400 questions and responses from developers seeking help with 7 cryptography libraries on Stack Overflow; analyzing them to identify 16 underlying usability issues and studying see how prevalent they were across the 3 cryptography libraries for which we had the most questions for on Stack Overflow. Mapping our usability issues to Green and Smith's usability principles we identify 4 *usability smells* where the principles are not being observed. We suggest what developers may struggle the most with in the cryptography libraries, and where significant usability gains may be had for developers working to make libraries more usable.

## 1 Introduction

Cryptographic APIs are hard to use. Other work has developed recommendations, guidelines and principles for how to make them more usable—but how can we tell when such usability recommendations, guidelines and principles are not being implemented? In this paper we focus on the ten principles proposed by Green and Smith [14] (reproduced in Figure 1). We investigate two key questions: (i) what are the issues that developers face when using seven cryptography libraries and (ii) what are the telltale signs that one of the ten usability principles is being violated?

*Code smells* are indicators that a piece of software code may be of lower quality than desired [12]. A code smell signifies that, while a piece of code may not be broken, it is violating a design principle and may be fragile and prone to failure. For example, Fowler defines the *Shotgun Surgery* smell as:

> "You whiff this when every time you make a kind of change, you have to make a lot of little changes to a lot of different classes. When the changes are all over the place, they are hard to find, and it's easy to miss an important change." [12]

Code that smells of shotgun surgery may be correct and pass all the tests, but the smell suggests that there may be a deeper issue with the code's structure.

Following the idea of a code smell, a *usability smell* is an indicator that an interface may be difficult to use for its intended users. Past work has focused on usability smells in graphical user interfaces (GUIs)—indicators that end users may struggle to use an application [2, 16]. However, usability issues are not limited to GUIs. Developers struggle with programming interfaces in the same way that users struggle with user interfaces. For example, past work has suggested that improving the quality of documentation would lead to developers needing to ask fewer questions about how to use libraries [19]. If the developer is unfamiliar with the library they will rely on the documentation provided with the library and their own programming knowledge to implement the required cryptographic tasks. If we look at a developer question and answer site, such as Stack Overflow (a popular developer question and answer help website), we might expect to see fewer questions asking for help with the basic usage of a library if it has improved its API documentation. However, as our analysis shows, these smells are present across the cryptography libraries we examined and all can make usability improvements to help developers use them successfully.

In order to identify developers' struggles with cryptographic libraries, we analyze 2,491 Stack Overflow questions. We examine questions about seven cryptographic libraries (Table 1), selected for their popularity and to encompass a broad range of languages and use-cases. We conduct a thematic analysis [5, 8, 23] of the questions and answers looking for the underlying reason the question was asked—be that because of missing documentation, confusion around an API, lack of cryptographic knowledge, or developers preferring Stack Overflow to other resources. We identify 16 thematic issues across our corpus of questions and measure their prevalence across the different libraries. We relate these issues

| Library | URL |
|---|---|
| OpenSSL | https://github.com/openssl/openssl |
| NaCl | http://nacl.cr.yp.to |
| libsodium | https://github.com/jedisct1/libsodium |
| Bouncy Castle | https://bouncycastle.org/java.html |
| SJCL | https://github.com/bitwiseshiftleft/sjcl |
| Crypto-JS | https://github.com/brix/crypto-js |
| PyCrypto | https://www.dlitz.net/software/pycrypto/ |

Table 1: Cryptography libraries examined in this paper.

back to Green and Smith's usability principles and identify *four usability smells* that indicate that *specific* principles are not being implemented fully. Finally we make suggestions, based on the prevalence of smells in each of the libraries, as to how library developers can better implement the principles to reduce the smells and make their API more usable.

The novel contributions of our investigation are as follows:

- An empirical validation of Green and Smith's principles showing when a principle is not being applied but also identifying issues that Green and Smith's principles currently do not capture.
- The thematic analysis of 2,491 Stack Overflow questions to assess the usability of cryptographic libraries.
- Identification of 16 thematic issues across 7 cryptographic libraries—capturing developers' struggles with regards to the usability of these libraries codified into four usability smells (Needs a super sleuth, Confusion reigns, Needs a post mortem, and Doesn't play well with others) which are signs that particular Green and Smith principles are not being fully implemented by a given library; before giving an overview of the prevalence of theses 16 issues, and 4 smells in 3 of the libraries (those that had over a hundred questions with a score ≥ 2).

## 2 Background and related work

The background and related work falls into two broad categories: research on usability issues of APIs in general; and work focusing on such issues in cryptography and security libraries. We discuss each of these bodies of work next.

### 2.1 Usability issues of APIs

Many studies have addressed the usability of APIs and why they can be difficult to learn and use. Zibran et al. [27] reviewed 1513 bug posts across five different repositories to identify the API usability issues that were reflected in the bug posts by the developers who used the APIs. They found 22 different API usability factors. We adopt a similar approach

**Abstract** Integrate cryptographic functionality into standard APIs so regular developers do not have to interact with cryptographic APIs in the first place.

**Powerful** Sufficiently powerful to satisfy both security and non-security requirements.

**Comprehensible** Easy to learn, even without cryptographic expertise.

**Ergonomic** Don't break the developer's paradigm.

**Intuitive** Easy to use, even without documentation.

**Failing** Hard to misuse. Incorrect use should lead to visible errors.

**Safe** Defaults should be safe and never ambiguous.

**Testable** Testing mode. If developers need to run tests they can reduce the security for convenience.

**Readable** Easy to read and maintain code that uses it/Updatability.

**Explained** Assist with/handle end-user interaction, and provide error messages where possible.

Figure 1: Green and Smith's 10 usable cryptography API principles, reproduced from [14]. We have given each principle a short name to allow easy reference throughout the paper.

and review the questions developers have about each one of our selected cryptographic libraries and see what prevalent usability issues arise. However, we investigate further to identify the usability smells from each library that contribute to the violation of the Green and Smith principles.

Other work has explored ways to measure the usability of existing APIs. Scheller and Kúhn proposed a framework for measuring an API's usability against a set of usability aspects [24]. Dekel and Herbsleb noted that many APIs place notes about when it is appropriate to use certain functions [9]. They developed a tool (eMoose) to integrate these notes into developer's editors (when using an annotated API) and found that developers who used their tool debugged programs quicker than those who only had access to the documentation. In follow up work [10] they noted that the key behind eMoose's success was not that eMoose made the notes immediately available, but rather that it helped provide a *scent* for programmers trying to debug their code—a hint that there was something that *could* go wrong and that prompted them to further read the documentation. Our work identifies *usability smells* that library developers and maintainers can use to understand how they may improve the usability of their libraries in line with the Green and Smith principles.

Helping developers avoid mistakes has been studied extensively with many papers suggesting ways APIs can be improved to help avoid mistakes and to speed up debugging when they inevitably occur. Bloch asserted general principles for API design that would produce a usable API [4]. These principles were summarized as 39 different maxims, though Bloch noted that good API design is a craft and couldn't be entirely captured by lists of rules. Others have proposed similar lists of metrics, often developed from studying or surveying developers. Ko and Yann studied the way developers use APIs [18]. They found that developers do not only need detailed worked examples but also good explanations of the concepts, parameters and ideas behind the API's design. Piccioni et al. [22] ran a study to assess the usability of an API by comparing the programmer's expectations to their performance. They found that issues with naming convention and types confused programmers, poor documentation made programming harder and that overly flexible APIs confused less experienced programmers. Clarke and Becker adapted the *cognitive dimensions* framework, used to describe the usability of user interfaces [15], to evaluate the usability of a class's API [6]. They suggested a list of ten dimensions that APIs should be judged on based on the original cognitive dimensions framework and two new ones that capture how much work any individual operation does. Our work complements such research by identifying the signs and smells – based on questions that developers ask when seeking help—that suggest developers are struggling to use an API.

As developers have changed so have their sources of documentation. Stack Overflow is increasingly used as a primary source of documentation. Parnin et al. surveyed questions about three popular (non-cryptographic) APIs on the site. They found that, on average, 80% of the API functions would be covered by at least one Stack Overflow question, but that only a relatively small pool of *experts* answered them [21]. Treude and Robillard looked at ways to extract insight from Stack Overflow questions and then how to integrate them with developer's toolchains [25]. In a small study of developers they established that developers found these extra insights helpful when programming. In a similar manner, our work studies developers' struggles through an analysis of the questions they ask on Stack Overflow. Our work adds to the literature by using Stack Overflow not to gain insight into developers, but rather into the usability issues of APIs the developers use.

## 2.2 Usability issues of cryptography and security libraries

Nadi et al. [20] examined over 1,000 Java cryptography-related questions and developed a set of 5 *obstacles* capturing the developer's problem based on the top 100 questions. Whilst Nadi et al. just looked at Java developers' struggles our work is broader examining developers' struggles with libraries for multiple languages and systems. They reported a low inter-rater reliability score (kappa = 0.41). We also relate these back to principles to identify underlying issues in the form of usability smells.

Egele et al. looked at the use of cryptographic APIs in Android applications [11]. They found mistakes in 88% of the apps that used cryptographic APIs. They developed a static analysis tool to identify mistakes automatically and proposed three usability guidelines to help developers avoid making mistakes in the future. Mindermann et al. studied the usability of cryptography APIs for the Rust programming language [19]. They noted that, whilst insecure defaults (and defaults in general) do not occur frequently in cryptographic libraries, very few projects warn about depreciated cryptography techniques or encourage developers to use more secure methods. They produced a list of 13 recommendations for cryptography APIs.

Various studies have assessed the usability of specific cryptographic libraries, e.g., [4, 14, 19] and developed a set of usability metrics. For example, one common guideline is:

> "Use a prominent location to link to the documentation, e.g., at the start page of the repository." [19]

Another guideline suggests providing examples so that developers can see how to use the library:

> "Example code should be exemplary. If an API is used widely, its examples will be the archetypes for thousands of programs." [4]

The issues we identify in this paper complement these guidelines by acting as usability smells—usability principles tell us how to make libraries more usable, the smells suggest where users are struggling due to such principles not being observed or implemented.

Studies have also shown that even if a cryptographic library is powerful developers may suggest to use an alternative cryptographic library which, although more usable, may be poorly implemented [1, 14]. For instance, Acar et al. [1] showed that the Python cryptographic library *Keyczar*, despite claiming to be designed for usability, was challenging to use because of poor documentation and lack of documented support for the key generation task. Surveying the developers after their study Acar et al. found that developers frequently found issues with missing documentation and examples in Python cryptographic libraries. We also find in our analysis of Stack Overflow questions that many developers struggle and ask questions due to issues with *missing documentation* and a need for *example code*, alongside several other issues.

If not all cryptography libraries are equally usable, then what issues do developers struggle with when using them? The analysis presented in this paper sheds light on such issues and identifies the usability smells that indicate that usability principles are not being fully observed in the design of a particular library.

## 3 Method

We investigate:

1. what issues with cryptographic libraries cause developers to seek help and ask questions on the Stack Overflow question and answer site;
2. how prevalent are the usability issues that we identify in seven cryptography libraries; and
3. what are the usability smells that are indicative of failures to implement Green and Smith's usability principles, and their prevalence in the 3 libraries for which we have sufficient data.

To answer these questions we selected seven cryptographic libraries to examine (Table 1), based on their prominence as well as to include a breadth of languages—C, Java, JavaScript and Python—and use-cases. Other libraries exist, but these seven cover a representative sample of what various cryptographic libraries currently look like including those with many users (OpenSSL) to those with only a few (SJCL). Additionally, two of the libraries (NaCl and libsodium) describe themselves as being *usable*, so we would hope to see a range of issues and differences between the usable cryptographic libraries and the not-so usable ones.

We scraped 2491 questions from Stack Overflow, using the library names as search terms and selecting only questions with a score greater than 1 to help avoid low-value and badly worded questions. Stack Overflow uses a reputation system to help combat spam—users with sufficient reputation[1] are allowed to vote for the usefulness of a question, which becomes the question's *score*.

We conducted a manual review of all 2491 questions: identifying the underlying issue, capturing common themes between questions, and verifying the validity of the answer. In order to do so, we studied the full description of the question on Stack Overflow to help us pinpoint the core theme of the issue. We also analyzed the explanations provided by other developers in the answer section to that question. The questions in our corpus cover a wide time period. Therefore, to address the issue of validity, once we studied the question and any related answers, we reviewed the prominent link of the cryptographic library to assess whether the question was still valid and unaddressed by the library. For example, if the developer said that they could not find documentation for a specific feature they wished to use, we checked if the documentation was still unavailable in the current version of resources for the cryptographic library. Questions that did not relate to an issue with the library itself, for example, due to users not understanding the behavior of their operating system's dynamic linker (we return to this issue in the Discussion; Section 7) or where the developer mistakenly attributed their question to a library, were not considered further. In total, we analyzed 2317 relevant questions.

We used thematic analysis, a qualitative research method used to extract themes from text [5, 8, 23], to identify recurring themes such as the need for documentation, build and compatibility issues within the Stack Overflow questions. We developed our themes by iteratively labelling questions, and then reviewing and discussing the labelling. We arrived at a set of 16 themes that captured the different issues developer's faced and ascribed a final, single theme to each of the questions we examined. Initially we used multiple themes, however we found only 4 cases where a question had multiple labels, so we simplified and ascribed the theme that best categorized the underlying reason the question was asked. We repeated the labelling with a regularly selected 10% subset of the questions analyzed by a second researcher and calculated Cohen's kappa—a commonly used measure of inter-rater agreement [7]. Cohen's kappa was 0.76 indicating that our coding was consistent between mappers.

### 3.1 Threats to validity

The questions in our corpus cover a period of several years. There is a danger that as time and the cryptographic libraries themselves change that the issues developers face could also change. To mitigate this we validated that usability issue identified in the library were still present in the current version. For example, if we attributed an issue to the documentation being missing, we validated that we still couldn't find the relevant documentation.

There is also the danger that an issue faced by a developer may be due to a particular problem faced solely by that developer and not a more general problem. To mitigate against this we selected questions which had a score greater than one—that is to say that more users of Stack Overflow believed the question to be worthwhile than not. Stack Overflow's reputation system is designed to help remove questions that have already been answered, and those that are of low-value (for example, questions where a developer has not asked a question, or questions where students are attempting to have their coursework answered for them). By selecting only questions with a positive score we help avoid some noise.

During our thematic analysis, each question was mapped to a single theme, with the dominant theme being picked in the case that a question could be attributed to multiple themes. For the most part, however, questions could be ascribed to a single theme and multiple themes were rare so a 1–1 mapping was used for consistency.

To identify the usability smells, we map the issues we identify to the usability principles that library developers should be implementing as identified by Green and Smith [14]. Various others have suggested different principles for developers (as we discuss in Section 2). We selected Green and Smith's principles because their principles have not currently been

[1] At the time of writing: 15 reputation to vote up, 125 to vote down.

validated, and were themselves a synthesis of other usability research [4] focused on usability and security issues. Other principles could be validated using the same methods and corpus as we have used however, and our dataset is available for comparative studies.

## 4 What usability issues do developers face?

Our thematic analysis reveals 16 usability issues with which developers struggle (Figure 2) categorized into 7 themes as shown in Figure 3. We discuss each of the issues and give some examples to demonstrate how they manifest in the questions posed by the developers.

### 4.1 Missing information

***Missing Documentation.*** A developer states that they wish to use a function or form a feature that has components supported by the library but cannot find relevant information in the library documentation:

> "So I already know how to specify locations for trusted certificates using `SSL_CTX_load_verify_locations()`.[⋯] But nothing is mentioned about the trusted system certificates residing in the OPENSSLDIR."

***Looking for Example Code.*** Not all library functionality needs an example, but it can be helpful to document common use-cases. The developer wishes to use a function supported by the library and requests examples of how the function is used. In the question the developer may address the quality of the example code or lack thereof:

> "I'm attempting to run:
>
> ```
> openssl pkcs12 -export -in "path.p12"
> -out "newfile.pem"
> ```
>
> but I get an error.
>
> ```
> unable to load private key
> ```
>
> How do I extract the certificate in PEM from PKCS#12 store using OpenSSL?"

This differs from *passing the buck* in that the developer has identified the functionality they want to use and made attempt at solving it. They have stated the problem they want to solve and have asked for a example in order to debug their own attempt.

***Clarity of documentation.*** The developer found the documentation or output but found it vague or unclear in describing what exactly it does:

> "How can I interpret openssl speed output?
>
> I ran openssl speed on my Ubuntu computer. [⋯] what is 'Doing md4 for 3s' mean? does it mean do the whole test for 3 times/seconds? what does '1809773 md4's in 2.99s' mean? what does '8192 size blocks' mean? [⋯] And the above, last lines of openssl speed md4 output - what does they mean exactly?"

### 4.2 Not knowing what to do.

***Passing the buck.*** The developer delegates their question to the Stack Overflow community, even though a quick search for the issue on the library website returns the answer needed:

> "I'm trying to convert the .cer file to .pem through openssl, the command is:
>
> ```
> openssl x509 -inform der -in
> certnew.cer -out ymcert.pem
> ```
>
> and that's the errors I'm getting:
>
> ```
> unable to load certificate
> ```
>
> What am I doing wrong?"

Rather than find the answer themselves the developer has *passed the buck* and used Stack Overflow to get the answer rather than search existing resources, as reflected in the response:

> "[⋯] like explained by ssl.com, a .cer file [⋯]"

Passing the buck differs from other issues, such as *what's gone wrong here*, in that the developer has made no attempt to solve the problem. They have encountered a problem and want someone else to give them the answer rather than work it out or find an existing solution by themselves.

***Lack of knowledge.*** There were many instances where new users struggled with the functions provided by a library due to the lack of knowledge they had about the concepts of cryptography. For example:

> "[⋯] I'm using OpenSSL to avoid pay for it. I created my certificate this way: [⋯]
>
> But when I navigate to the website I get an "error" telling me that this is an "Untrusted certificate": The security certificate presented by this website was not issued by a trusted certificate authority."

This lack of knowledge is implicitly highlighted in the answer:

> "What you get from OpenSSL tool is a self signed certificate. Of course it is not trusted by any browser, as who can say you are worth the trust.
>
> Please buy a certificate if you want to set up a public web site [⋯]"

**Missing Documentation.** The cryptographic library does not have documentation available to address the issue.

**Example Code.** The developer asks for code examples to learn how to use a specific feature of the library or to learn how to implement some behavior. An example is either missing or lacking somehow.

**Clarity of Documentation.** The library has documentation for the developer's issue, but it is unclear or lacking additional information. The developer asks for clarification.

**Passing the buck.** The developer asks Stack Overflow, even though documentation regarding their issue has been given. Also questions where they ask a simple question which they answer themselves.

**Lack of Knowledge.** The developer does not have foundation level cryptography knowledge. The developer is new to cryptography as a subject and, in turn, the features of the cryptographic library.

**Unsupported Feature.** The crypto library does not support a security feature the developer wants to implement.

**Borrowed Mental Models.** The developer requests a mapping of a functionality between cryptographic libraries.

**Abstraction Issue.** Issues addressing the level of abstraction provided in the code of the cryptographic library. The developer wants a more detailed explanation than is provided by the documentation.

**What's gone wrong here?** The developer has code that looks like it should work, but fails—they are looking for an explanation why.

**API Misuse.** The developer has incorrectly used a specific feature from the cryptographic library.

**Should I use this?** The developer says what they wish to implement and asks which methods would be most apt to use.

**How should I use this?** The developer does not understand how to correctly use a feature or its various parameters.

**Build Issues.** Issues related to the setup of the cryptographic library and running provided tests.

**Performance Issues.** Issues regarding the performance of the cryptographic library.

**Compatibility Issues.** Issues related to integrating features from the cryptographic library with other libraries and tools.

**Deprecated Feature.** Issues addressing that a specific feature is not working, later to conclude that the feature is deprecated.

Figure 2: The 16 issues identified through a thematic Analysis of Stack Overflow Questions.

## 4.3 Not knowing if it can do

***Unsupported feature.*** The developer wants to do something that the library does not support. This may suggest that the library is unclear about what it can and cannot do:

> "Has anybody Implemented ElGamal using OpenSSL or even inside?"

***Borrowed mental models.*** The developer is trying to take a mental model about how one library works and apply it to different one:

> "How to recreate the following signing cmd-line OpenSSL call using M2Crypto in Python?:
>
> This works perfectly in command-line, I would like to do the same using M2Crypto in Python code.
>
> [⋯]"

The developer has tried to apply concepts from one library to another and has become confused when that doesn't work. This differs from *passing-the-buck* in that they are not unwilling to learn, they just don't know that the concepts differ. Passing-the-Buck is where a developer doesn't know how to use a library and tries to get someone else to tell them. They don't care about learning and just want to be told what to do.

## 4.4 Programming is hard

***Abstraction issue.*** The developer needs help with an abstraction provided by the library. They've seen the documentation but they lack knowledge of the underlying abstraction to understand it. They need more help:

> I am trying to get my head around public key encryption using the openssl implementation of rsa in

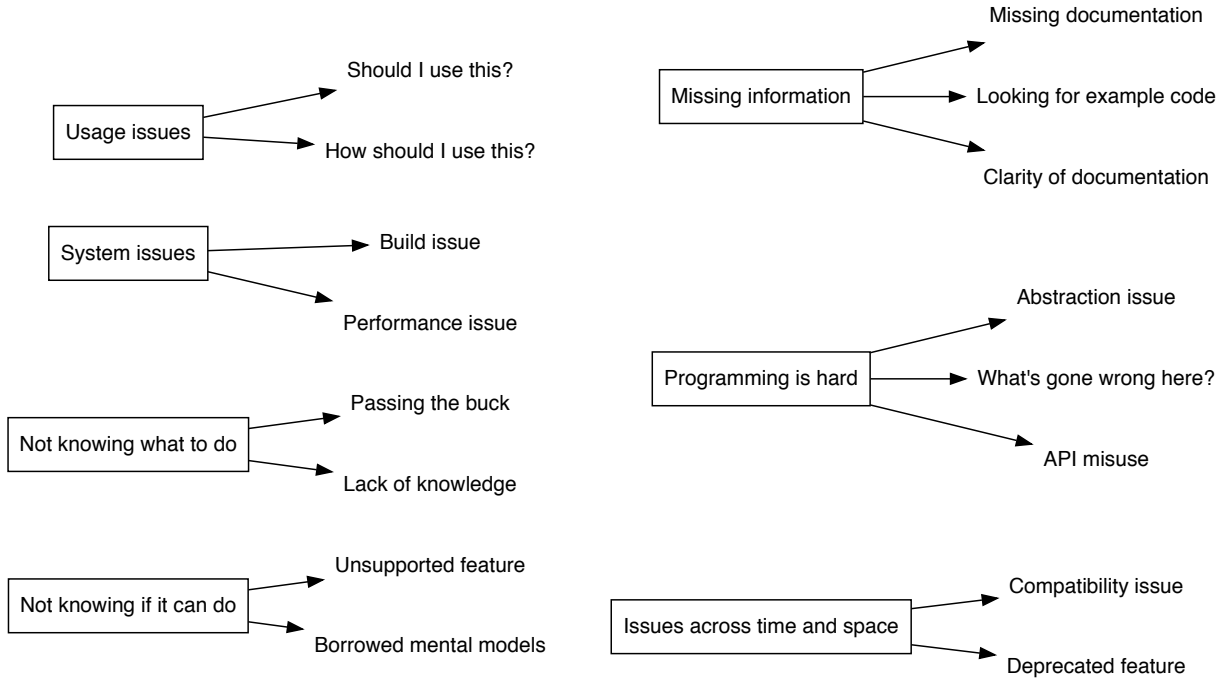



Figure 3: Categorization of the 16 issues identified through the thematic analysis.

> C++. Can you help? So far these are my thoughts (please do correct if necessary) [···] I see these two functions: [···] If Alice is to generate *rsa, how does this yield the rsa key pair? Is there something like rsa_public and rsa_private which are derived from rsa? Does *rsa contain both public and private key and the above function automatically strips out the necessary key depending on whether it requires the public or private part? [···]

***What's gone wrong here?*** The developer has tried to use the library but has failed. They have given a specific example and asked Stack Overflow to suggest what has gone wrong:

> "Here is a certificate in x509 format that stores the public key and the modulo:
>
> ```
> const unsigned char
> *certificateDataBytes = /*data*/;
> ```
>
> Using OpenSSL and C, how can I convert it into an RSA object? I've tried several methods but I can't get it to work in `RSA_public_encrypt`"

***API misuse.*** API Misuse represents questions where the developer incorrectly uses a function and they are corrected by another developer, usually supported with an explanation of the answer. For example:

> "[···] I'm trying to build a handshake protocol for my own project and am having issues with the server converting the clients RSA's public key to a Bignum. It works in my client code, but the server segfaults when attempting to convert the hex value of the clients public RSA to a bignum."

In the response to the question, the correct use of the function is explained:

> "RSA new() only creates the RSA struct, it does not create any of the bignum objects inside that struct, like the n and e fields. [···]"

## 4.5 Usage issues.

***Should I use this?*** Developers have tasks and features in mind for which they want to know whether they should use a specific library function or not—or if there are two or more functions, which one should they use? In other cases, developers want to know whether the choices they make regarding the security of their application are appropriate:

> "I'm trying to build two functions using PyCrypto that accept two parameters: the message and the key, and then encrypt/decrypt the message.
>
> I found several links on the web to help me out, but each one of them has flaws:
>
> [···] Also, there are several modes, which one is recommended? I don't know what to use :/"

This differs from *missing documentation* where the developer is searching for specific API documentation, in that here they are unsure about which part of the API they want to use in the first place.

***How should I use this?*** In contrast with *Should I use this*, in such cases the developer knows what they want to use, but is confused about some of the parameters involved:

> "How to compute RSA-SHA1(sha1WithRSAEncryption) value with OpenSSL?"

## 4.6 System issues

***Build issues.*** To use an API developers must first build it (and run tests). This causes problems for the developer:

> "Error compiling OpenSSL with MinGW/MSYS"

Or

> "How to build OpenSSL to generate libcrypto.a with Android NDK and Windows"

***Performance issues.*** The developer wants to use a library but finds that it isn't performant enough for their use-case. They seek help in optimizing their use of the library:

> "[···] profiling has revealed [···] 40% of my library runtime is devoted to creating and taking down HMAC_CTX's behind the scenes. [···] How do I get rid of the 40% overhead on each invocation in a (1) thread-safe / (2) resume-able state way? [···]"

## 4.7 Issues across space and time

***Compatibility issues.*** The developer is struggling to integrate the library in question with another platform or library. For instance, out of the 2022 questions pertaining to OpenSSL, 244 were related to compatibility issues:

> "Encrypt in C# using OpenSSL compatible format, decrypt in Poco:
>
> I'm trying to encrypt (aes-128-cbc) in Win OS using a OpenSSL compatible format and decrypt on Linux OS using Poco::Crypto that is a wrapper of OpenSSL. "

***Deprecated feature.*** The developer is trying to do something the library once supported, but doesn't know that the latest version has deprecated it:

> "After a few days of scouring the internet and openssl docs i've hit a wall [···]."

In the answers the developer realizes that they are using an outdated API.

> "Thanks to JWW and indiv i was able to solve my problem, it was an issue with me using older API's, and improper return checking. Solution: [···]:"

| Issue | OpenSSL | Libsodium | NaCl | Bouncy Castle | SJCL | Crypto-JS | PyCrypto |
|---|---|---|---|---|---|---|---|
| Missing Documentation | 256 (13%) | 3 (9%) | 5 (12%) | 31 (17%) | 4 (27%) | 2 (6%) | 7 (4%) |
| Example Code | 128 (6%) | | 1 (2%) | 10 (5%) | 2 (13%) | | 4 (3%) |
| Clarity of Documentation | 92 (5%) | | 3 (7%) | 2 (1%) | | | 6 (4%) |
| Passing the buck | 136 (7%) | 2 (6%) | 4 (10%) | 22 (12%) | 4 (27%) | 17 (49%) | 10 (6%) |
| Lack of Knowledge | 44 (2%) | 6 (19%) | 3 (7%) | 19 (10%) | | 4 (11%) | 17 (11%) |
| Unsupported Feature | 24 (1%) | | 1 (2%) | 5 (3%) | | | 7 (4%) |
| Borrowed Mental Models | 56 (3%) | | 2 (5%) | 1 (1%) | | | |
| Abstraction Issue | 40 (2%) | | 2 (5%) | 2 (1%) | 2 (13%) | 2 (6%) | 10 (6%) |
| What's gone wrong here? | 259 (13%) | 1 (3%) | 2 (5%) | 24 (13%) | | 3 (9?%) | 16 (10%) |
| API Misuse | 11 (1%) | | 1 (2%) | 6 (3%) | | | 7 (4%) |
| Should I use this? | 84 (4%) | | 8 (19%) | 19 (10%) | | 1 (3%) | 8 (5%) |
| How should I use this? | 80 (4%) | | | 10 (5%) | | 2 (6%) | 4 (3%) |
| Build Issue | 362 (18%) | 7 (22%) | 3 (7%) | 15 (8%) | | 3 (9%) | 57 (36%) |
| Performance Issue | 20 (1%) | | 1 (2%) | | | | |
| Compatibility Issue | 244 (12%) | 7 (22%) | 6 (14%) | 8 (4%) | 3 (20%) | 1 (3%) | 5 (3%) |
| Deprecated Feature | 20 (1%) | | | 9 (5%) | | | 1 (1%) |
| Not Relevant | 166 (8%) | 6 (19%) | | 2 (1%) | | | |

Table 2: Count of the number of Stack Overflow Questions attributed to each usability issue per library. Zero counts ommited.

# 5 How widespread are the issues across the seven libraries?

Table 2 shows the number of times each issue appeared during our thematic analysis for each cryptographic library; and suggests common issues across the libraries. *Missing Documentation* is a common issue: it suggests that developers face an issue in the first stages of using a cryptographic library as they are unable to locate documentation to support them. For instance, SJCL provides the code of each of its functions as its only developer support resource, and so can be made much stronger if they considered adding documentation to support the functions provided.

*Passing the Buck* and *Lack of Knowledge* highlight issues associated with developer behaviors instead of the cryptographic libraries themselves. Passing the Buck issues are common showing that developers have a tendency to pose questions on Stack Overflow, while the resources addressing the very questions are provided by the library and easy to locate. Many instances are recorded under the OpenSSL library, along with Bouncy Castle and Crypto-JS. Bouncy Castle and PyCrypto have a high percentage of questions associated with Lack of Knowledge. Developers address their lack of knowledge in their questions and request support in learning

cryptographic concepts in order to use functions from these libraries.

There are many occasions where the developer has a specific feature in mind for a project and wants to know how to securely implement this feature into the project. The reason the number of questions defined under *How should I use this?* is high for OpenSSL, for example, may be because the developer believes that other developers have already implemented the feature they had in mind. So the developer resorts to finding the specific implementation on developer community sites such as Stack Overflow instead of building their feature using the documentation provided by the cryptographic library. This could also explain why there are many questions where developers show an example of their broken code and request guidance with debugging. The answers usually come in the form of task-based examples of how to correctly implement, something to which the developers respond well.

Other than Missing Documentation, developers also highlight difficulties they have while setting up OpenSSL and running the provided tests. Reviewing the questions, we see that developers have projects in mind and intend to implement OpenSSL with other platforms they are using. This raises many questions associated with *Compatibility*. Developers find it very difficult to integrate OpenSSL with other platforms, a particularly pertinent issue as OpenSSL is widely used—and for large-scale projects. However, having compatibility issues makes OpenSSL less usable as developers cannot easily reconcile implementation of security requirements with other requirements for their projects.

# 6 Usability smells

Having identified the above issues, we map them on to Green and Smith's 10 principles (shown in Figure 1) in order to identify the usability smells that are indicative of one or more of the principles not being fully observed. The purpose of these smells is not to identify usability issues with a library early, but rather to guide work to improve a library's usability based on where developers appear to struggle between releases of a library as part of the software lifecycle. We note that Green and Smith's principles are written in a positive manner—for example:

> "Integrate cryptographic functionality into standard APIs so regular developers do not have to interact with cryptographic APIs in the first place."

In contrast, the issues we identify from the thematic study are written in a negative context—for example:

> "Missing Documentation: The cryptographic library does not have documentation available to address the issue."

The two viewpoints however are linked—if a library developer fails to fully implement a usability principle, then we might expect to see questions indicating that the library users are struggling with one of the usability issues we identify. Our mapping between usability principles and usability issues is presented in Table 3. For each issue we identified, we considered whether it would indicate failing to implement one of Green and Smith's principles. We did not map the *lack of knowledge* or *passing the buck* issues as these are attributable to specific developer behaviors and do not represent failures to implement usability within a library, and so do not map to Green and Smith's principles. For example the *borrowed mental model* issue is present when a developer expects one library to work similarly to another; this is mapped to the *ergonomic* principle as it indicates a failure to not break the developer's paradigm.

Based on this mapping, we identify four usability smells. In the same fashion as Fowler [12], we describe them as *whiffs*.

## 6.1 Needs a super sleuth

*Issues at play:* Missing documentation; Example code; Clarity of documentation.

You whiff this when documentation is missing, unclear or there is a lack of example code pertaining to how to use the library. The information to achieve the task you are intending to undertake is hard to find or understand in a way that can make the library work for your needs easily. You need to be a super-sleuth to find the documentation and decipher its meaning!

By not breaking the developer's paradigm (the ergonomic principle), developers can intuitively use the library with fewer references to the documentation or example code. By providing visible and early errors (the failing principle) developers can quickly understand when something is wrong and fix it themselves.

## 6.2 Confusion reigns

*Issues at play:* Should I use this; How should I use this; Abstraction issue; Borrowed mental models.

You can catch a whiff of this when developers are designing and prototyping their programs—they are trying to decide whether this is the right library to use and how to start using it. They are unclear as to how to use the library, perhaps having confused some concepts or borrowed a mental model they have for another library that isn't relevant here.

By making the library easy to use even without documentation (intuitive principle) a developer can quickly work out if they should use a library, how to use it and understand the abstraction it provides. If it is easy to learn (comprehensible principle) they can quickly evaluate it. If it uses standard APIs (abstraction principle) they can quickly figure out its use without worrying about details. By not breaking the developer's paradigm (ergonomic principle) they can reuse existing

| Whiff | Issue | Abstract | Powerful | Comprehensible | Ergonomic | Intuitive | Failing | Safe | Testable | Readable | Explained |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Need a super-sleuth | Missing Documentation | | | | ● | | | | | | |
| | Example code | | | | ● | | ● | | | | |
| | Clarity of documentation | | | | ● | | ● | | | | |
| Confusion reigns | Should I use this? | | | ● | | ● | | | | | |
| | How should I use this? | ● | ● | | | ● | | | | | |
| | Abstraction issues | ● | | | | ● | | | | | |
| | Borrowed mental models | | | | ● | | | | | | |
| Needs a post-mortem | What's gone wrong here? | | | | ● | ● | | | | ● | ● |
| | Unsupported feature | | | | | ● | | | | | |
| | API misuse | | | | | | ● | ● | | | |
| | Deprecated feature | | | | | | ● | | | | |
| Doesn't play well with others | Build issues | | | | | | | | ● | | |
| | Compatibility issues | | ● | | | | | | | | |
| | Performance issues | | | | | | | | | | |

Table 3: Mapping between developer issues and Green & Smith principles.

mental models about how similar libraries behave.

## 6.3 Needs a post-mortem

*Issues at play:* What's gone wrong here; Unsupported feature; API misuse; Deprecated feature.

If the *confusion reigns* whiff concerns the smells pre-coding, then this whiff occurs after they have written some code. The developer has used the library but something has gone wrong. Either they have used the library incorrectly or they are struggling to work out if it is an issue with the library itself. Perhaps an update to the library has broken their code, or led them to believe that it can do something it can't—either way the code needs a post-mortem.

By not breaking the developers model (ergonomic principle) developers can quickly guess whether their code is erroneous and work out what's gone wrong. If it is easy to use (intuitive principle) then this aids with debugging what's gone wrong as well as figuring out the capabilities and features of the library. Being hard to misuse (failing principle) avoids API misuse, and prevents API designers from deprecating APIs without a warning. If the defaults are safe and sensible (safe principle), then developers may avoid the complex API features and their potential misuses. Making the code easy to read (readable principle) means that if a developer needs to dive into the source to figure a bug out, they can do so with the minimum of fuss. Finally by helping developers with end-user interaction (explained principle), library designers can ensure developers do things in a standard way hence avoiding the need to ask if other people have done things the same way or differently.

## 6.4 Doesn't play well with others

*Issues at play:* Build issue; Compatibility issue; Performance issue.

If a library is going to be easy to use, developers have to be able to use it in the first place. This smell occurs when the library won't build, won't integrate with other libraries and build systems, and is a resource hog without providing a clear explanation why.

This smell doesn't appear to be particularly well covered by Green and Smith's issues. By adding a testing mode with only a subset of the features active (testable principle) developers can avoid having to build all the dependencies and get the library built for early testing and prototyping. By making the library powerful enough to satisfy security and non-security requirements (powerful principle) developers can more easily integrate it with other less flexible libraries.

We group performance issues under this smell, however we could not see any of Green and Smith's principles that exactly covered this usability aspect. In describing the *explained* principle, Green and Smith suggest:

> "Firstly, most developers using a security API do not have a firm grasp on the cryptographic or security background and thus would be hard pressed to explain to the end-user what went wrong" [14]

Perhaps by extending this principle to include not just *why things go wrong*, but also *why things take so long* this additional issue could be covered by Green and Smith's ten principles. Alternatively the *powerful* principle could be extended to cover not just the developer's primary security and non-security functionality requirements, but also cover the performance aspects.

| Whiff | Issue | OpenSSL | Bouncy Castle | PyCrypto |
|---|---|---|---|---|
| Needs a super sleuth | *Whiffiness factor* | 10% ● | 11% ● | 4% ● |
| | Missing documentation | 13% | 17% | 4% |
| | Example code | 6% | 5% | 3% |
| | Clarity of documentation | 5% | 1% | 4% |
| Confusion reigns | *Whiffiness factor* | 3% ● | 6% ● | 4% ● |
| | Should I use this? | 4% | 10% | 5% |
| | How should I use this? | 4% | 5% | 3% |
| | Abstraction Issue | 2% | 10% | 6% |
| | Borrowed mental models | 3% | 1% | 0% |
| Needs a post mortem | *Whiffiness factor* | 10% ● | 11% ● | 8% ● |
| | What's gone wrong here? | 12% | 13% | 10% |
| | Unsupported feature | 1% | 3% | 4% |
| | API misuse | 1% | 3% | 4% |
| | Deprecated feature | 1% | 5% | 1% |
| Doesn't play well with others | *Whiffiness factor* | 11% ● | 5% ● | 22% ● |
| | Build issue | 18% | 8% | 36% |
| | Compatibility issue | 12% | 4% | 3% |
| | Performance issue | 1% | 0% | 0% |

Table 4: What whiffs can you smell on each library? Percentages of the questions for each library that were mapped to each issue are shown, along side a *Whiffiness factor*, based on the weighted average, that indicates how strong the smell is:

●: particularly pungent (weighted average > 10%);

●: merest whiff (weighted average ≥ 2, ≤ 10%).

## 7 Discussion

With four whiffs established, Table 4 describes how smelly 3 of the crypto libraries appear to be—OpenSSL, Bouncy Castle and PyCrypto. For the remaining 4 libraries we lack a sufficient volume of questions to make any meaningful statement about the issues with which the library's users may struggle. However, for these 3 libraries we have 2,022, 185 and 160 questions respectively and so can consider where the *pain points* for developers using these libraries may lie. We include the libraries with fewer questions in our thematic analysis in order to reduce skew towards the issues prevalent in the more frequently queried libraries, however we lack the volume of questions required to suggest what the pain-points for developers are in the 4 remaining libraries. We do not claim that there is a fault in any of the libraries—rather we suggest what the most frequent issues that some developers struggle with when using them are—and where the biggest usability gains might be had. Future work should explore and find the underlying cause for the smell and establish *why* developers appear to be struggling.

For each library we add a *Whiffiness factor* (based on the weighted average of the percentage frequency of the issues associated with each whiff). All the libraries we looked at smell a little of *needing a super sleuth* with OpenSSL and Bouncy Castle users especially struggled with missing documentation. Despite this the overall whiffiness of this smell appeared to be low, as there were fewer questions over all 3 libraries associated with these issues—this suggests that documentation may be improving; and whilst documenting more of the library and giving more examples will help users, there may be bigger usability gains to be had elsewhere.

As for the *confusion reigns* whiff, again, the libraries all seem to show some signs of it—with the issue being particularly pronounced for Bouncy Castle, where we saw many developers asking whether it was appropriate to use this library, and having particular issues with the abstractions it provides. This is somewhat surprising as, at least for the Java version, Bouncy Castle integrates with the Java Cryptography Architecture which provides a standard API for libraries providing cryptography functionality. Bouncy Castle also provides its own API, and supports languages other than Java—perhaps offering too much choice confuses developer as to the parameters for a specific version. Focusing on the intuitive and comprehensible principles, i.e., by making the library easier to learn and understand without the need for expertise, should help reduce this smell.

The *doesn't play well with others* whiff was present for all three libraries. OpenSSL and PyCrypto in particular struggled with *build issues*, whereas Bouncy Castle (which is available as a precompiled JAR file) had fewer issues associated with building the library. Integrating software into systems is known to be difficult [13], but offering prebuilt images seems to go some way to mitigating this. Building the library is just the first step for OpenSSL however, as the library has to be linked into the final compiled program. When mapping the Stack Overflow questions, we saw several examples of developers asking about the dynamic linker. For instance:

> "I am using OpenSSL in my project.library is detecting but getting some errors like below:
>
> `Error:(23) undefined reference to 'RSA_generate_key' [⋯]`
>
> I included appropriate .so files in appropriate folder. I am not getting reason behind the undefined reference error.please help me to solve this issue."

These are not library usability issues as they represent a misunderstanding about the host-system and tools rather than the library itself. So we did not map them to an issue (the question in the specific example above was resolved by the developer updating their Makefile). The issue was common enough, however, that we believe that there may be a serious usability issue integrating libraries with systems, and a gap in the literature in looking into these issues. Further work is

needed to map out what these issues are, how common they are and what we can do to mitigate them.

The final whiff we identified, the *needs a post mortem* smell, was prevalent in the OpenSSL, Bouncy Castle and PyCrypto libraries. For these libraries the biggest contributing issue to this smell was that of developers trying to establish what had gone wrong with their programs. Making the code more readable and the libraries more intuitive to use even without documentation should help to make the debugging process easier and mitigate this smell. OpenSSL and Bouncy Castle are lower-level than other cryptography libraries providing greater access to their internals and crypto primitives. For these libraries we would expect an increase in the number of questions by developers trying to debug the code, simply because they wrap things up less into high-level APIs and offer more scope for developers to make a mistake. Perhaps then it is not unreasonable to expect lower-level libraries to display this issue more than the higher-level ones.

OpenSSL in particular, has been criticised in the past for being hard to use [1, 17, 26]. Kamp in particular argued for someone to:

> "Please Put OpenSSL Out of Its Misery.
>
> OpenSSL must die, for it will never get any better." [17]

Our analysis certainly suggests that OpenSSL is a bit stinky—in particular it seems that developers struggle a lot debugging it and in finding the documentation. Ignoring those issues, however, it is similar to the other crypto libraries and sometimes a little bit better (it seems better at abstraction, describing its parameters, for example)—at the very least both Bouncy Castle and PyCrypto appear harder to debug. OpenSSL gets a lot of stick for being unusable but perhaps it doesn't deserve it all—it has a general pong of poor usability but there are other libraries with sharper, more specific, stenches out there too.

## 8 Conclusion

How can we tell what a developer is struggling with when using a crypto library? Through our analysis of a substantial corpus (2491) of questions from Stack Overflow, we found 16 issues and four whiffs that suggest when developers are struggling. By linking these smells to the usability principles by Green and Smith [14], we can suggest how to improve crypto libraries and make them more usable for developers. Our study offers evidence to validate parts of Green and Smith's heuristics, but also highlights issues that were missed. Their usability principles suggest ways to mitigate most of the issues we identify; however issues associated with the *doesn't play well with others* smell (in particular *build* and *performance* issues) suggest the need for an additional principle to help cover these issues.

Our whiffs capture the general problems developers have when using crypto libraries. Not all libraries smell the same, and improvements to *usable* crypto libraries appear to be paying off with fewer usability smells. By smelling carefully we can find the pain point for developers and help improve usability. Libraries will perhaps always be a bit smelly given the challenges of catering for the requirements of a wide and diverse set of developers and applications; but by integrating usability principles we can at least make them less so.

## 9 Acknowledgements

This work is supported by funding from the National Cyber Security Centre and in part by the Engineering and Physical Sciences Research Council grant EP/P011799/2: Why Johnny doesn't write secure software.

## References

[1] Y. Acar, M. Backes, S. Fahl, S. Garfinkel, D. Kim, M. L. Mazurek, and C. Stransky. Comparing the usability of cryptographic APIs. In *2017 IEEE Symposium on Security and Privacy*, pages 154–171, May 2017.

[2] D. Almeida, J. C. Campos, J. Saraiva, and J. C. Silva. Towards a catalog of usability smells. In *Proceedings of the 30th Annual ACM Symposium on Applied Computing*, pages 175–181. ACM, 2015.

[3] D. G. Altman. *Practical statistics for medical research*. CRC press, 1990.

[4] J. Bloch. How to design a good API and why it matters. In *Companion to the 21st ACM SIGPLAN symposium on Object-oriented programming systems, languages, and applications*, pages 506–507. ACM, 2006.

[5] V. Braun and V. Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.

[6] S. Clarke and C. Becker. Using the cognitive dimensions framework to evaluate the usability of a class library. In *Proceedings of the First Joint Conference of EASE PPIG (PPIG 15)*, 2003.

[7] J. Cohen. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46, 1960.

[8] J. W. Creswell and J. D. Creswell. *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications, 1994.

[9] U. Dekel and J. D. Herbsleb. Improving api documentation usability with knowledge pushing. In *Proceedings*

*of the 31st International Conference on Software Engineering*, pages 320–330. IEEE Computer Society, 2009.

[10] U. Dekel and J. D. Herbsleb. Reading the documentation of invoked API functions in program comprehension. In *2009 IEEE 17th International Conference on Program Comprehension (ICPC 2009)*, pages 168–177. IEEE, 2009.

[11] M. Egele, D. Brumley, Y. Fratantonio, and C. Kruegel. An empirical study of cryptographic misuse in android applications. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, pages 73–84. ACM, 2013.

[12] M. Fowler. *Refactoring: improving the design of existing code*. Addison-Wesley Professional, 1999.

[13] D. Garlan, R. Allen, and J. Ockerbloom. Architectural mismatch or why it's hard to build systems out of existing parts. In *1995 17th International Conference on Software Engineering*, pages 179–179. IEEE, 1995.

[14] M. Green and M. Smith. Developers are not the enemy!: The need for usable security APIs. *IEEE Security & Privacy*, 14(5):40–46, 2016.

[15] T. R. G. Green and M. Petre. Usability analysis of visual programming environments: A 'cognitive dimensions' framework. *Journal of visual languages and computing*, 7(2):131–174, 1996.

[16] P. Harms and J. Grabowski. Usage-based automatic detection of usability smells. In *International Conference on Human-Centred Software Engineering*, pages 217–234. Springer, 2014.

[17] P. H. Kamp. Please put OpenSSL out of its misery. *ACM Queue*, 12(3):20–23, 2014.

[18] A. J. Ko and Y. Riche. The role of conceptual knowledge in API usability. In *Visual Languages and Human-Centric Computing (VL/HCC), 2011 IEEE Symposium on*, pages 173–176. IEEE, 2011.

[19] K. Mindermann, P. Keck, and S. Wagner. How usable are Rust cryptography APIs? *arXiv preprint arXiv:1806.04929*, 2018.

[20] S. Nadi, S. Kriüger, M. Mezini, and E. Bodden. Jumping through hoops: Why do Java developers struggle with cryptography APIs? In *2016 IEEE/ACM 38th International Conference on Software Engineering (ICSE)*, pages 935–946, 2016.

[21] C. Parnin, C. Treude, L. Grammel, and M. A. Storey. Crowd documentation: Exploring the coverage and the dynamics of API discussions on stack overflow. *Georgia Institute of Technology, Tech. Rep*, 2012.

[22] M. Piccioni, C. A. Furia, and B. Meyer. An empirical study of API usability. In *Empirical Software Engineering and Measurement, 2013 ACM/IEEE international symposium on*, pages 5–14. IEEE, 2013.

[23] J. Saldaña. *The coding manual for qualitative researchers*. Sage, 2015.

[24] T. Scheller and E. Kühn. Automated measurement of api usability: The api concepts framework. *Information and Software Technology*, 61:145–162, 2015.

[25] C. Treude and M. P. Robillard. Augmenting API documentation with insights from stack overflow. In *Software Engineering (ICSE), 2016 IEEE/ACM 38th International Conference on*, pages 392–403. IEEE, 2016.

[26] M. Ukrop and V. Matyas. Why Johnny the developer can't work with public key certificates. In *Cryptographers' Track at the RSA Conference*, pages 45–64. Springer, 2018.

[27] M. F. Zibran, F. Z. Eishita, and C. K. Roy. Useful, but usable? factors affecting the usability of APIs. In *2011 18th Working Conference on Reverse Engineering*, pages 151–155. IEEE, 2011.

# System Administrators Prefer Command Line Interfaces, Don't They? An Exploratory Study of Firewall Interfaces

Artem Voronkov
*Karlstad University*

Leonardo A. Martucci
*Karlstad University*

Stefan Lindskog
*Karlstad University*

## Abstract

A graphical user interface (GUI) represents the most common option for interacting with computer systems. However, according to the literature system administrators often favor command line interfaces (CLIs). The goal of our work is to investigate which interfaces system administrators prefer, and which they actually utilize in their daily tasks. We collected experiences and opinions from 300 system administrators with the help of an online survey. All our respondents are system administrators, who work or have worked with firewalls. Our results show that only 32% of the respondents prefer CLIs for managing firewalls, while the corresponding figure is 60% for GUIs. We report the mentioned strengths and limitations of each interface and the tasks for which they are utilized by the system administrators. Based on these results, we provide design recommendations for firewall interfaces.

## 1 Introduction

Firewalls are systems designed to regulate network traffic, and are often the first line of defense in computer networks. The maintenance and configuration of firewalls is the responsibility of system administrators. System administrators have multiple methods available to interact with firewalls, e.g. via a command line interface (CLI), graphical user interface (GUI), or application programming interface (API). Although visualization offers an effective approach to exploring and managing data, the use of GUIs by system administrators is not taken for granted. According to the literature, the main instrument for system administrators is the CLI [2, 9, 18].


*USENIX Symposium on Usable Privacy and Security (SOUPS)* 2019.
August 11–13, 2019, Santa Clara, CA, USA.

In this paper, we examine how system administrators interact with firewalls. The goal of our study is to gain a better understanding of the following questions:

Q1: What firewall interfaces do system administrators use?

Q2: What firewall interfaces do they prefer?

Additionally, we want to gain insights into which of the interfaces are beneficial for which tasks, and what strengths and limitations they have. To answer our research questions, we surveyed 300 system administrators and collected their experiences and opinions of utilized firewall interfaces through an online survey.

Unexpectedly, our results show that 70% of the system administrators work primarily with firewall GUIs, with 60% preferring GUIs as a main instrument. The system administrators mainly choose GUIs because they provide better visual representations of data, are easier to create and modify rules with, and are convenient for occasional use. Relatively few system administrators utilize a CLI as their primary or preferred firewall interface: 24% and 32%, respectively. According to our respondents, the main reasons for choosing command line interfaces are their flexibility, efficiency of use, superior functionality, and performance; aspects in which GUIs are deficient.

The contributions of our work are summarized as follows:

- We conduct an online study on the preferences of system administrators regarding firewall interfaces, with 300 volunteer participants.
- Using the gathered data, we classify and report the main strengths and limitations of CLIs and GUIs.
- We provide insights into tasks in which utilizing a CLI or GUI is advantageous for system administrators.
- We provide some recommendations for designers and developers of firewall interfaces, taking into account the main problems of the two interfaces.

The remainder of this paper presents a review of work related to our study in Section 2, describes our research methodology in Section 3, and presents the results in Section 4. A discussion of the findings, limitations, and our design recommendations is presented in Section 5. Finally, concluding remarks are provided in Section 6.

## 2 Related Work

Despite the fact that GUIs are known to be convenient for the presentation of large amounts of information, their use is limited in the field of system configuration, as noted by Mahendiran et al. [10].

Botta et al. [2] and Haber and Bailey [9] reported the results of two independent ethnographic studies describing the routines and activities of system administrators. Haber and Bailey followed the daily work of three system administrators, and reported their preference of CLIs over GUIs owing to their speed, scalability, reliability, transparency and trustworthiness. These findings are in line with the interviews of Botta et al., involving a dozen IT professionals who reported being more comfortable with CLIs than GUIs, especially because of their versatility. Botta et al. also highlighted the reliability problem of GUIs that *"write configuration files that sometimes do not take effect"* and *"write unnecessary, noisy markup into configuration files."*

For a study with 101 participants, Takayama and Kandogan [18] reported that 65% of the participants were primarily CLI users, because CLIs are considered to be more reliable, fast, robust, trustworthy, and accurate. Furthermore, the authors pointed out that trust is critical in the adoption of a technology.

However, system administrators require graphical tools that can facilitate their daily work and make it less error-prone [10]. This is especially relevant for security system administrators, as their work has been demonstrated to be more complex [6].

Recent research has sought to leverage the benefits of information visualization in designing interfaces for network security. Shiravi et al. [15] presented a survey of visualization systems in network security in general, while Voronkov et al. [21] reviewed papers specifically concerning firewalls. The authors of both papers identified limitations of existing visualization techniques and suggested future research directions.

Xu et al. [22] argued that *"system configuration becomes a new human–computer interaction (HCI) problem,"* and that *"classic interface design principles are not sufficient for system configuration."* A variety of research studies [9, 19, 20] have attempted to address these problems and suggest appropriate design principles for system configuration.

Although interface preferences of system administrators have been studied in the literature, the present work represents the first large-scale study investigating firewall interfaces, with 300 participants. Furthermore, we aim to investigate whether there have been changes in preferences, as it has been over 10 years since the studies of Botta et al. [2], Haber and Bailey [9], and Takayama and Kandogan [18] were published. Another important aspect of our work is the qualitative analysis of participants' comments regarding the strengths and limitations of firewall interfaces, as well as tasks in which these interfaces are superior.

## 3 Methodology

We collected both quantitative and qualitative data on the interactions between system administrators and firewall interfaces through an online survey ($N = 300$). In this section, the methodology and demographics of the participants are described, while the remainder of the quantitative data and qualitative results are presented in Section 4.

### 3.1 Survey Details

We collected the data through an online survey, which ran for six weeks from April to June 2018.[1] The survey utilized skip logic (also known as branch logic or conditional branching) and consisted of up to 14 questions, four of which were open-ended. The close-ended questions required an answer and we also encouraged the participants to answer the open-ended questions, although these were not mandatory.

The survey consisted of two parts. In the first, we asked the participants about the following aspects of their interactions with firewalls:

- How much time on average they spend working with firewalls.
- Which firewall interface they mainly work with, and which interface they prefer.
- Which tasks are easier with which firewall interface.
- What strengths and limitations those interfaces have.

Only general questions about firewall interfaces were asked in the survey. No questions about specific vendor solutions were included. In the second part of the survey, demographically related questions were asked, such as on age, gender, and expertise.

We kept the survey short to minimize respondent fatigue. The survey took an average of 177 seconds ($SD = 106$, $M = 148$, $Q1 = 101$, and $Q3 = 228$ seconds) of the participants' time to be answered.

Prior to dissemination, the survey was pre-tested with six users. Based on their feedback, a few questions were slightly altered to eliminate some ambiguity in the wording, although no significant changes were necessary. For wider coverage, the survey was translated from the original (English) language into three others (Portuguese, Russian, and Swedish) by bilingual speakers.

[1] The survey is available at https://www.soscisurvey.de/firewall_interfaces/

## 3.2 Recruitment and Participants

The participants for the study were recruited using various channels:

1. System administrators' forums. The "Sysadmin" subreddit yielded the majority of our participants.[2] Another contributor was the SysAdmins.ru forum.[3]

2. System administrators' mailing lists. We contacted several system administrators from our professional networks and asked them to distribute the survey via system administrator mailing lists of which they are members.

Of 516 participants that started our online survey, 303 completed it (ca. 59% completion rate). After the quality check, three participants were removed as they filled out nonsensical answers. Table 1 summarizes the demographics of the remaining 300 participants. Our sample is heavily skewed owing to specificity of the target audience (the percentage of female system administrators is known to be very low [1]) and recruitment method. A majority of the participants (approximately 80%) were recruited via the "Sysadmin" subreddit, which led to the sample being more male (only 7.5% of the subreddit members are female [3]) and younger than the general population, owing to the demographics of Reddit users [14]. All participants were volunteers, and no financial compensation was offered.

## 3.3 Survey Data Analysis

The data were analyzed using a content analysis approach. With this approach, it is possible to analyze data qualitatively at the same time as quantifying it [8].

Two of the authors worked independently and coded participants' responses to the open-ended questions using an initial (open) coding approach [13]. Two coding procedures were performed: one before and one after the final codebook. We utilized NVivo for all coding.[4] NVivo helped us to organize and analyze the qualitative data, i.e. open-ended survey responses. NVivo provides methods to automatically or manually code the data. We used manual coding only, which comprises three approaches: 1) select and code content, 2) drag and drop selected content, and 3) in vivo coding.

After the authors completed the first coding procedure, they met, discussed their codes, consolidated them, and formed a final codebook, which consisted of 230 codes (see Section 6). Using the final codebook during the second coding procedure, 1570 coding references were identified. It is worth mentioning that each answer from a participant can have several different codes associated with it, but at most one instance of a single code.

[2] https://www.reddit.com/r/sysadmin/
[3] https://sysadmins.ru/
[4] https://www.qsrinternational.com/nvivo/home

Table 1: Participant demographics ($N = 300$).

| | Metric | Participants |
|---|---|---|
| **Age** | 18-24 | 34 (11.3%) |
| | 25-34 | 142 (47.3%) |
| | 35-44 | 86 (28.7%) |
| | 45-54 | 25 (8.3%) |
| | 55-64 | 9 (3.0%) |
| | Prefer not to answer | 4 (1.3%) |
| **Gender** | Female | 3 (1.0%) |
| | Male | 285 (95.0%) |
| | Other | 1 (0.3%) |
| | Prefer not to answer | 11 (3.7%) |
| **Time per week (on average) spent on managing firewalls** | <1 hour/week | 106 (35.3%) |
| | 1-4 hours/week | 117 (39.0%) |
| | 5-8 hours/week | 35 (11.7%) |
| | 9-12 hours/week | 11 (3.7%) |
| | 13+ hours/week | 21 (7.0%) |
| | Do not directly manage firewalls | 10 (3.3%) |
| **Experience as system administrator** | <1 year | 6 (2.0%) |
| | 1-3 years | 46 (15.3%) |
| | 4-6 years | 64 (21.3%) |
| | 7-9 years | 39 (13.0%) |
| | 10+ years | 145 (48.3%) |
| **Proficiency with firewalls** | Basic knowledge | 20 (6.7%) |
| | Intermediate | 114 (38.0%) |
| | Advanced | 114 (38.0%) |
| | Expert | 52 (17.3%) |
| **Language** | English | 256 (85.3%) |
| | Portuguese | 7 (2.3%) |
| | Russian | 21 (7.0%) |
| | Swedish | 16 (5.3%) |

The Cohen's kappa inter-rater reliability value for the final codes was 0.79, indicating an excellent agreement between the coders [4]. The cases in which the coders varied in the final codes were resolved by the first author, who examined respondents' answers and assigned the most appropriate code.

## 3.4 Ethical Considerations

The survey was conducted in accordance with the Swedish Ethical Review Act [16] and the Good Research Practice guidelines from the Swedish Research Council [17]. No sensitive personal data were collected and no mental or physical interventions took place. Therefore, no explicit ethical approval was required for this study. The following precautions were taken into consideration to ensure that the participants were treated ethically and with respect:

- The participants provided informed consent before starting the survey. The informed consent form stated the purpose of the study, its approximate duration, our commitment to confidentiality, and their rights as participants,

including the right to withdraw from the study at any point in time.

- Only (the minimal) necessary personal data (see Table 1) were collected.
- No sensitive personal data were collected.

# 4 Results

We describe the survey results by providing both quantitative and qualitative data in Sections 4.1–4.2. In Section 4.3 we report on the suitability of firewall CLIs and GUIs for different tasks.

## 4.1 Quantitative Data

Seventy percent of the participants in our survey are primarily firewall GUI users, and 60% prefer GUIs to text-based interfaces when having to deal with a firewall (see Table 2). Approximately a quarter of the polled system administrators primarily work with textual interfaces (24% for CLI and 2% for API), and slightly over one third prefer to use these as their main interface: 32% and 4% for CLIs and APIs, respectively. The option `Other` indicates system administrators that use either a combination of the aforementioned interfaces or another type of firewall interface.

Based on our data, there may be a connection between a system administrator's proficiency with firewalls and the interface that they prefer to utilize. Table 3 shows that the stronger the firewall expertise of respondents, the lower the likelihood of utilizing GUIs. Seventy percent of the system administrators with a basic knowledge of firewalls prefer GUIs to any other interface, while this holds true for only 54% of firewall experts.

## 4.2 Qualitative Data

The thoughts and opinions of the system administrators received during our online survey were coded and grouped according to the following principles: 1) the type of interface: CLI, GUI, API, or other; and 2) the type of comment: positive, negative, or neutral. For the convenience of presenting the strengths and limitations of the interfaces, we categorized the codes as follows:

- We began classifying our codes according to the 10 usability heuristics introduced by Nielsen [12] (see Table 4).
- Because not all codes concerned usability, some of them did not fall into any of the 10 categories, and were further classified according to the ISO/IEC 25010 [5], a standard that defines systems and software quality models (see Figure 1). This includes aspects that are not covered by Nielsen's usability heuristics, such as security and reliability. Regarding usability, the ISO standard comprises appropriateness recognizability, learnability, operability, and accessibility, aspects that are not covered by Nielsen's usability heuristics.
- All remaining codes fell within the `Other` category.

Because the number of respondents who work with APIs or other interfaces is relatively small, we do not report the corresponding results in this paper.

The strengths and limitations of CLIs and GUIs (see Figures 2–5) are discussed in further detail in Sections 4.2.1–4.2.4. In each subsection, we examine the categories that cover 80% of all coding references, starting from the most popular. Note that subsections have different total numbers of coding references, and not all codes in each category are discussed in detail. For convenience, codes are highlighted in bold.

### 4.2.1 CLI Strengths

According to our respondents, CLIs have a number of strengths (the total number of coding references is 319):

1. Flexibility and efficiency of use (106 coding references; 33.2%). Several respondents (64 coding references) noted the possibility of **automation** as a strength of CLIs: *"CLIs are good targets for automation, even if the only thing you can do is bash scripting."* **User efficiency** was mentioned 42 times. One respondent stated: *"CLIs have a high signal-to-noise ratio, and are therefore preferable to everything else."*

2. Functional suitability (62 coding references; 19.4%). The **superior functionality** of CLIs was mentioned 37 times by system administrators: *"100% coverage of all firewall functionality supported by the OS kernel, unlike GUIs and APIs."* Other useful features of the interface, such as the **ability to work offline** and **ease of search**, were stated 16 times.

3. Usability (30 coding references; 9.4%). According to 12 system administrators, the user has **full control** with a firewall CLI: *"I do not see any reasonable way to be sure a firewall is doing the right thing without using a CLI."* Seven other respondents stated the advantages of managing a firewall with a CLI: *"Properly used, CLI is by far the best method to manage any system."*

4. Performance efficiency (22 coding references; 6.9%). The system administrators noted the superior **speed of operation** of CLIs (22 coding references), commenting *"[CLI] uses zero system resources"* and *"it is faster and does not take five minutes to load."*

Table 2: Relations between primary and preferred firewall interfaces based on the answers from our survey.

| | | Preferred interface | | | | Total |
|---|---|---|---|---|---|---|
| | | CLI | GUI | API | Other | |
| **Primary interface** | CLI | 61 | 7 | 3 | 2 | 73 (24.3%) |
| | GUI | 30 | 169 | 4 | 6 | 209 (69.7%) |
| | API | 0 | 1 | 4 | 0 | 5 (1.7%) |
| | Other | 4 | 2 | 0 | 7 | 13 (4.3%) |
| **Total** | | 95 (31.7%) | 179 (59.7%) | 11 (3.6%) | 15 (5.0%) | 300 (100.0%) |

Table 3: Relations between firewall proficiency and preferred firewall interfaces based on the answers from our survey.

| | | Preferred interface | | | | Total |
|---|---|---|---|---|---|---|
| | | CLI | GUI | API | Other | |
| **Proficiency** | Basic knowledge | 5 | 14 | 0 | 1 | 20 (6.7%) |
| | Intermediate | 30 | 72 | 2 | 10 | 114 (38%) |
| | Advanced | 43 | 65 | 5 | 1 | 114 (38%) |
| | Expert | 17 | 28 | 4 | 3 | 52 (17.3%) |
| **Total** | | 95 (31.7%) | 179 (59.7%) | 11 (3.6%) | 15 (5.0%) | 300 (100.0%) |

5. Visibility of system status (21 coding references; 6.6%). **Transparency** was mentioned 21 times as an important positive characteristic of CLIs: *"With a CLI, you know exactly what the firewall is doing."*

6. Reliability (16 coding references; 5.0%). Our respondents highlighted some strengths of CLIs, such as: **reliability**: *"... there is a lower incidence of random issues with the UI"*; **high availability**: *"I can do the same task via an SSH connection or even a KVM if the whole network is down. I can do that via a smartphone if I must."*; and ease of **configuration backup**: *"Backing up and restoring configurations easily through text files."*

#### 4.2.2 CLI Limitations

The main CLI limitations noted by our respondents are the following (the total number of coding references is 86):

1. Match between system and real world (22 coding references; 25.6%). The main problem, which was referenced 19 times, is a **long learning curve**. System administrators shared that *"CLI may be scary/overwhelming for a beginner/untrained user"* and *"There is typically a slightly higher learning curve associated with CLI, which can often be discouraging to unexperienced users."*

2. Usability (22 coding references; 25.6%). There are two codes that were referenced more than any others: CLIs are **not easy to use** (8 times) and **inconvenient data representation** (7 times). Two respondents stated: *"The CLI is not capable of representing all the firewall rule data in a clean and easy-to-read format"* and *"CLIs are terrible at generating visual information that is comprehensible by non-experts..."* Regarding the ease of use, one system administrator wrote that *"ease of use is a definite issue [of CLI]."*

3. Recognition rather than recall (10 coding references; 11.7%). The facts that CLIs are **less intuitive** and **less educational** were mentioned seven and three times, respectively. CLIs *"may be less intuitive than other interfaces"* and *"you cannot click your way around it in an attempt to* figure it out*."*

4. Error prevention (8 coding references; 9.3%). CLIs are **prone to errors**, both typographical and logical, and that fact was named 8 times by the respondents. One system administrator wrote that it is *"much easier to cause catastrophic failure quickly and effectively"* with a CLI.

5. Functional suitability (8 coding references; 9.3%). The absence of some auxiliary functionality was noted by eight respondents. A CLI *"has no Ctrl+F [searching] feature."*

#### 4.2.3 GUI Strengths

GUIs have several strengths (the total number of coding references is 586):

1. Usability (236 coding references; 40.3%). In general, GUIs are known to be user-friendly. Visual representations of data provide a **better understanding and/or overview of configuration** according to 124 coding references. One respondent shared with us that *"it [GUI] allows me to have a better understanding of a firewall's configuration while having that information displayed in a more organized manner when compared to a CLI."* The system administrators also stated that GUIs are **easy**

Table 4: Nielsen's usability heuristics [12].

| **Heuristics** | **Short explanation** |
|---|---|
| Visibility of system status | The system should always keep users informed about what is going on. |
| Match between system and real world | The system should speak the users' language. Information should appear in a natural and logical order. |
| User control and freedom | Users need clearly marked emergency exits. The system should support undo and redo. |
| Consistency and standards | Users should know whether different words, situations, or actions mean the same thing. The system should follow platform conventions. |
| Error prevention | The system should eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action. |
| Recognition rather than recall | The system should minimize the user's memory load. Instructions for use of the system should be visible or easily retrievable whenever appropriate. |
| Flexibility and efficiency of use | The system should have accelerators that can speed up interactions for expert users so that it can cater to both inexperienced and experienced users. |
| Aesthetic and minimalist design | Dialogues should not contain information that is irrelevant or rarely needed. |
| Help users recognize, diagnose, and recover from errors (we refer to this as assistance with errors) | The system should explain error messages in plain language, precisely indicate the problem, and constructively suggest a solution. |
| Help and documentation | Any system information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large. |

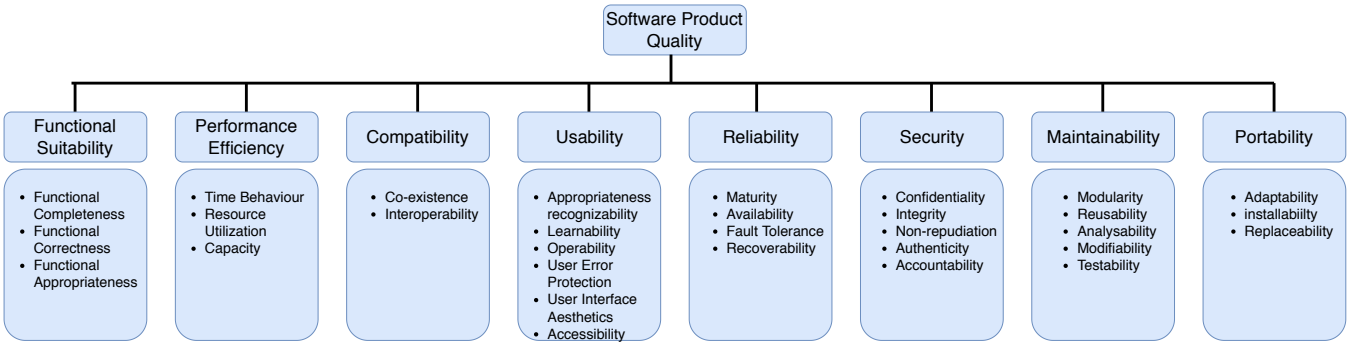



Figure 1: The software quality model ISO/IEC 25010 [5].

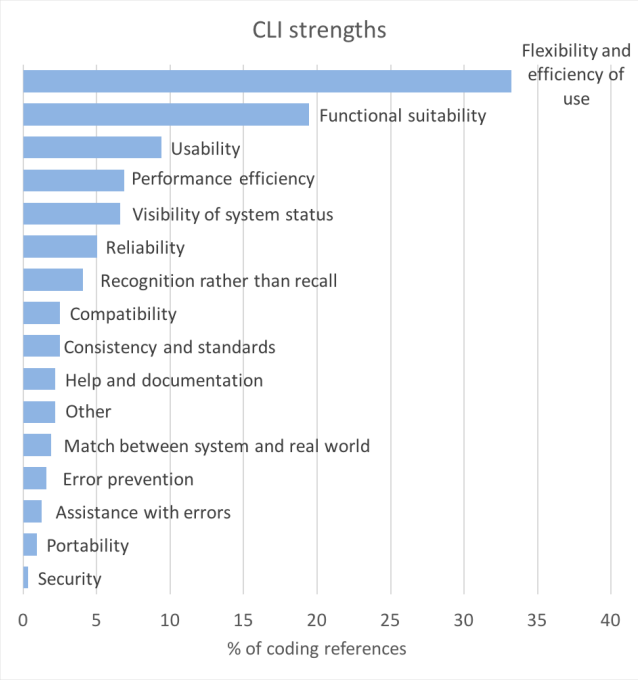



Figure 2: Classification of CLI strengths mentioned by our respondents. The total number of coding references is 319.

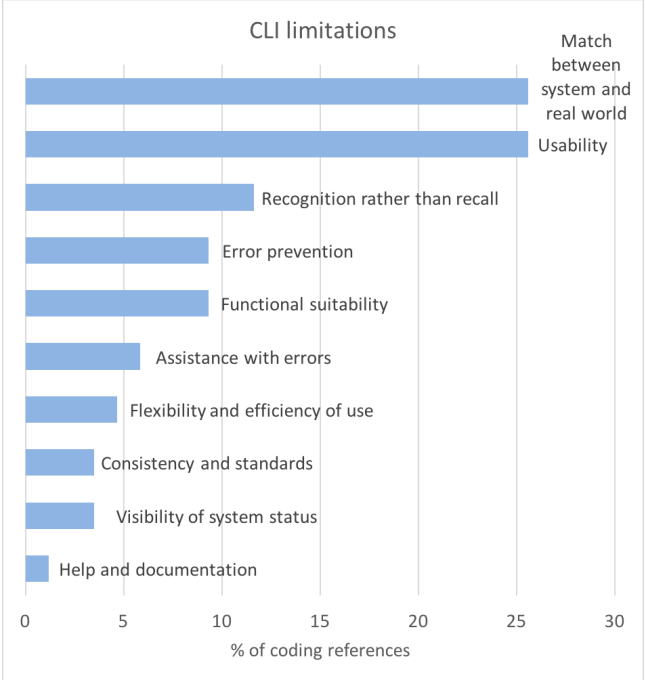



Figure 3: Classification of CLI limitations mentioned by our respondents. The total number of coding references is 86.

**to use** (49 times), **easy to manage and modify rules** with (19 times), good for **creating rules and policies** (16 times) and **good for people that struggle to work with text** (six times).

2. Functional suitability (120 coding references; 20.5%). The system administrators wrote that GUIs are excellent for a variety of tasks, such as **monitoring** (17 coding references), **reporting** (nine coding references), and **logging** (five coding references). Another strong aspect of GUIs is an **ease of displaying additional information** (20 coding references), such as graphs and statistics.

3. Recognition rather than recall (83 coding references; 14.2%). Being easy to navigate, GUIs are an irreplaceable tool that is **good for occasional use** (44 coding references). A system administrator shared: *"Because of my responsibilities as a general sysadmin [system administrator], management of the firewall takes up only a small part of my time, and having using the GUI for management means that I do not have to remember CLI commands."*

4. Flexibility and efficiency of use (45 coding references; 7.7%). **User efficiency** was named 20 times as a strength of GUIs: *"Makes it faster than using CLI to edit basic things on a firewall . . . ," "It just gives me a quicker and more visual grasp on what I am doing. Point, click, move on..."*

#### 4.2.4 GUI Limitations

Although the majority of our respondents prefer to use GUIs to other alternatives, several serious limitations of GUIs were named (the total number of coding references is 406):

1. Flexibility and efficiency of use (125 coding references; 30.8%). A **lack of automation** and **user inefficiency** in general when working with the interface were stated 102 times. This makes GUIs less useful for experts. One system administrator wrote: *"We are at a very bad time for GUI firewalls, because experts are the only ones who can effectively scale the workloads demanded of the modern IT infrastructure, and GUIs are almost useless for most experts in that regard."*

2. Functional suitability (56 coding references; 13.8%). According to 56 participants, the **reduced functionality** of GUIs is a serious issue: *"[GUI is] missing a lot of features/settings, so that you have to use CLI to make changes."*

3. Matching between the system and real world (38 coding references; 9.4%). Because a GUIs represents an **additional layer of abstraction**, the user may lack a deeper understanding of their actions (30 coding references). A system administrator formulated a drawback of GUIs as *"a lack of knowledge for the underlying system you are working on."* Another problem, named six times, is that GUIs may **generate less understandable configu-**

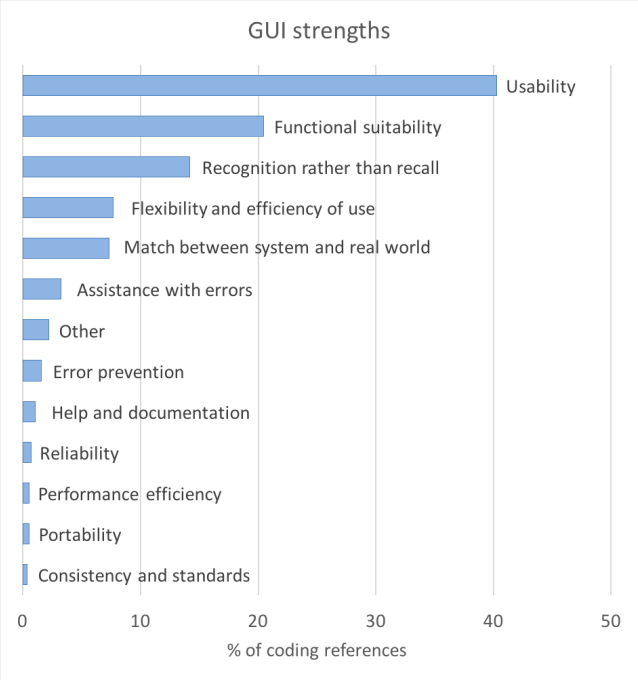



Figure 4: Classification of GUI strengths mentioned by our respondents. Total number of coding references is 586.

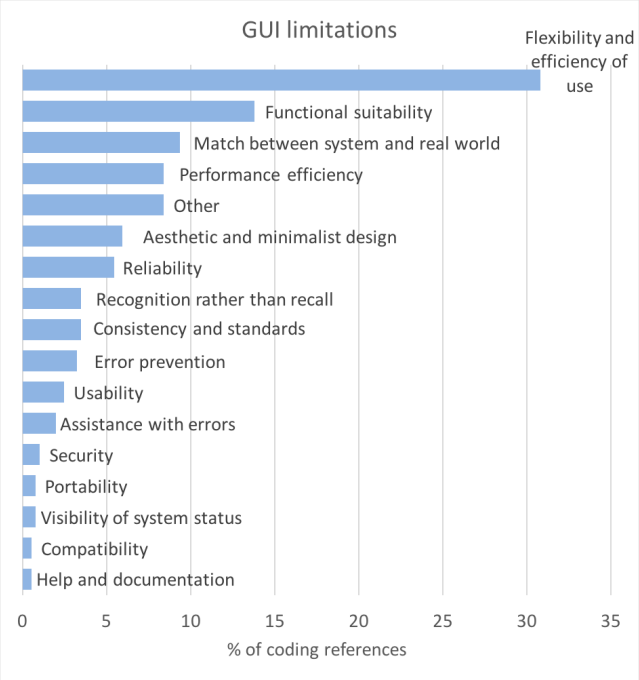



Figure 5: Classification of GUI limitations mentioned by our respondents. Total number of coding references is 406.

**ration files**: *"GUIs do not always generate configs that make logical/visual sense to a human."*

4. Performance efficiency (34 coding references; 8.4%). GUIs are highly demanding in terms of system resources, and for this reason are usually very **slow** (34 references): *"GUIs take more overhead to display and run, which may draw away from a firewall's processing power."*

5. Other (34 coding references; 8.4%). The system administrators stated a number of problems. The facts that GUIs **require additional equipment or software** and are **platform or browser dependent** were mentioned 12 times each. Two participants shared that *"A software client can be needed, which may not always be accessible..."* and *"Depending on browser it can be a horrible experience (slow, unresponsive, thus can cause issues with clicks being registered late or not at all)."* Several additional issues were mentioned by the system administrators, such as GUIs being **difficult to document** (five references): *"Unlike CLIs, documenting a GUI is mostly useless and defeats most of the purpose of a GUI,"* and **unavailable for particular firewalls** (three references): *"I currently do not have a firewall that supports GUI . . . ."*

6. Aesthetic and minimalist design (24 coding references; 5.9%). The respondents encountered **badly designed** GUIs (16 references): *"... some [GUIs] are horribly designed so it is hard to figure out how to do what you want to do."* Eight system administrators noted the problem of an **interface beauty and functionality trade-off**: *"Most [G]UIs are either poorly laid out making them difficult to use or are too user-friendly and do not have all settings available."*

7. Reliability (22 coding references; 5.4%). The respondents mentioned reliability issues (22 times) with GUIs: *"They can crash and become unresponsive, sometimes you cannot tell if it is processing a new config/update or if it is locked up."*

## 4.3 Suitability for Different Tasks

In addition to the strengths and limitations of firewall CLIs and GUIs, the system administrators informed us of which interface they deem to be the most suitable for each task.

All use cases associated with entering many similar rules on one computer or bulk changes on several computers at once require the use of a CLI. Furthermore, non-standard tasks in which, for example, rules with a set of advanced options are necessary, are more easily solved using a CLI according to the respondents. One system administrator stated:

> *"GUIs and APIs are terrible at handling special cases and rarely expose the underlying command structure properly (they always create a monopoly on how things are done). With CLIs, it is usually a straightforward process*

*to use the underlying commands and kernel modules directly (iptables and netfilter on Linux).*"

Firewall GUIs are preferable in more tasks according to the polled system administrators. The most frequently mentioned use cases in which the use of a GUI is beneficial are the creation of individual rules and building entire configurations. One respondent commented:

"*The more complicated a task is, the more important a GUI becomes. Who wants to set complex web proxy configuration options via CLI? Let us say that I want to proxy students, teachers, and administrators in a school differently. Let us start with just http request options for request headers, allowed auth[entication] methods, DLP scanning for HTTP POST, etc. Imagine the difference between clicking checkboxes and dropdown menus vs trying to type all this out via a CLI with some reference manual . . .*"

Another task that is easier to perform in a GUI is viewing and inspecting firewall rules and policies. GUIs usually have an option to link objects with rules, which allows the bigger picture to be observed.

"*Examining and working with firewall rules is an instructive example of where the CLI is not a good option. The CLI is not capable of representing all the firewall rule data in a clean and easy-to-read format. This is much easier on a GUI.*"

Monitoring is another use case in which the system administrators decide to use GUIs. GUIs provide the ability to view connection statistics, monitor traffic flows through real-time graphs, and so on, and are therefore preferred. The system administrators also tend to choose GUIs when there is a need to change the order of rules in a rule set.

## 5 Discussion

The collected quantitative data provides insights into the usage of different firewall interfaces that are considerably different from what has been previously published in the literature. We observe a significant shift towards the use of GUIs, although CLIs have been widely utilized by system administrators in the recent past [18]. There are three possible explanations for this shift.

The first concerns the case of security tools, in particular firewalls, where designers attempt to follow the design principles formulated for system administration tools. Our participants confirmed that firewall GUI implementations are improving. One system administrator opined:

"*Decades of GUI development: a 2D mouse and keyboard with a keyboard shortcuts interface instead of the serial text in/out of a classic CLI. More available and powerful searching, sorting, and filtering of information; discoverability of available commands; and visual/graphical possibilities of a large and high-res screen.*"

The second possible reason is that the number of system administrators has significantly increased, including those with limited technical expertise, as described by Xu et al. [22]. The statement on less experienced system administrators is not valid for our data sample, as 83% of the respondents have worked for over three years as system administrators (see Table 1).

The third possible reason is that there are many system administrators who are not security experts, but rather general purpose system administrators. As we can also observe from Table 1, 74% of the respondents spend no more than four hours per week managing firewalls. Therefore, they are most likely general purpose system administrators, and this explains their reluctance to work with firewall CLIs, which are less usable, require more learning time, and are prone to errors according to our participants.

Our qualitative data show that GUIs are less preferable for experts compared to system administrators with a lower firewall proficiency. The respondents noted that GUIs are not very useful for experts, as they severely restrict the user with limited functionality, a low operation speed, and low user interaction efficiency owing to the lack of automation capabilities.

Another feature that we noticed when analyzing the data from the survey is that our respondents' preferences for one interface do not always depend on their strengths or limitations. Sometimes system administrators are more comfortable with a CLI or GUI simply because they familiarized themselves with this interface first. One respondent stated:

"*I am old school and there were no GUIs back in the day, so it [CLI] is more comfortable for me.*"

Another possible reason is that system administrators do not always have experience with other interfaces, and therefore cannot objectively compare their strengths and limitations.

### 5.1 Limitations

One of the limitations of our study is that most of the respondents were recruited through online forums for system administrators. Because the survey participants were volunteers, there is a self-selection bias that leads to the sample not being fully representative.

Furthermore, the study was conducted online and we could not observe the participants answering the questions. Moreover, some of the answers were ambiguous, and so we had to interpret them, which could lead to a distortion of the meaning that the respondent had originally intended. For example, the comment *"slow"* can refer both to the speed of operation of the software and the speed of interaction between the user and

interface. There is also a possibility of questions being misunderstood or misinterpreted by the participants. Additionally, self-report surveys have several common limitations [7], such as social desirability biases and acquiescent responses.

We mitigated the limitations by carefully considering the design of the survey, pretesting it with several participants, making it anonymous so that people could answer honestly, and shortening it to minimize respondent fatigue.

## 5.2 Design Recommendations

Our survey identified some problems for both CLIs and GUIs that should be taken into account. In this section, we present some design recommendations for CLIs and GUIs based on the results of our survey, as well as discussing the benefits of combining these two interfaces into one.

As one respondent noted:

> "*CLI interfaces are not usually as forgiving as other interfaces. If you are not paying attention, then the slightest typo could cause large issues.*"

Our recommendation is to employ a syntactical verification of commands when a user types in an instruction to prevent errors in firewall configuration processes.

Furthermore, because CLIs have a reasonably long learning curve, assistance in writing rules is necessary for less experienced system administrators. Respondents noted the following:

> "*It may be difficult to compose rules [in CLIs] without an example.*"

Providing a knowledge base of examples of rules could be a useful approach.

We make three recommendations regarding GUIs. First, the system administrators complained about the speed of operation of GUIs. Our recommendation is to not make GUIs bloated, so that they do not consume a lot of system resources and can be run on mediocre hardware.

The second recommendation relates to the GUI installation process. As one of the system administrators commented:

> "*They [GUIs] are not really for beginners because of the initial setup required to configure them.*"

Because the highest percentage of GUI use is among system administrators with the least firewall expertise, installing a firewall should not be a complicated task.

In addition, to increase the speed of user interaction with a firewall GUI it is necessary to allow system administrators to create their own combinations of hotkeys for the most popular actions. This will help to make GUIs more attractive for firewall experts.

While we have provided recommendations for how to improve each interface, there remain problems that are difficult to solve within one interface. For example, textual interfaces are inherently inadequate for presenting a large amount of information:

> "*When a config file has over 2000+ lines it is easy to lose track of what is what [in CLI].*"

Another limitation originates from the concept of a CLI: it is impossible to create and edit rules using the mouse cursor and check boxes, which in some cases can significantly increase the productivity of a firewall operator. For GUIs, the problem is the lack of automation tools, as was noted by a large number of respondents.

A more effective solution would be to combine two interfaces into one, with the ability to seamlessly switch from one to the other, so that interacting with one interface affects the other. Such an approach can leverage the strengths of each interface while mitigating their limitations [11]. A GUI can provide an overview of configurations and display additional graphs and statistics, as well as being used to create rules, while a CLI can offer on-demand access to the powerful automation capabilities. Such a combined interface could be suitable for users with different firewall expertise. Less experienced system administrators could be trained to use the CLI by viewing the underlying text-based commands while working in the GUI. Expert users could continue using commands to create rules, while using the GUI for a better policy overview. We strongly believe that such a firewall interface would be widely accepted in the system administration community.

# 6 Conclusion

In this work, we present an online study concerning system administrators, in which we examine how they interact with different firewall interfaces. The survey results show that 70% of the polled system administrators are primarily GUI users, and 60% prefer this interface for interacting with a firewall. This finding differs from previously published findings in the literature, in which CLIs were claimed to be the first choice of system administrators.

We classify the strengths and limitations of firewall CLIs and GUIs. Our participants report that CLIs are flexible, efficient, transparent, reliable, and achieve ultimate functionality and a good performance. However, they are inconvenient for representing data, do not help users by preventing errors, and have a long learning curve. On the other hand, GUIs help users to perceive firewall configuration information more effectively and have a shorter learning curve compared to CLIs. They are also easy to use, easy to create and modify rules with, and good for occasional use. Regarding the limitations of GUIs, they restrict users with limited functionality, a low operation speed, and a low user interaction efficiency. They are neither transparent nor reliable. In addition, we report

the preferred interface for each task according to the system administrators.

Our findings present opportunities for future research. A well-designed firewall interface should predict and interpret its user's needs and assist them in becoming proficient with the firewall. In this case, the system administrator is satisfied with the firewall and can efficiently perform the required work. On the other hand, a poorly designed firewall interface might hinder the successful execution of tasks and lead to the future disuse of that solution. We provide some design recommendations that should be taken into account by designers aiming to develop better CLIs and GUIs.

## Acknowledgments

We are grateful to all the system administrators that participated in our study. We would also like to thank the moderators of the *Sysadmin* subreddit (`r/sysadmin`) for allowing us to reach out to their community.

This work was supported by the Knowledge Foundation of Sweden HITS project and by the Swedish Foundation for Strategic Research SURPRISE project.

## Availability

The final codes and other details are available at https://github.com/soups2019-126/supplementary_material.

## References

[1] Catherine Ashcraft, Brad McLain, and Elizabeth Eger. *Women in tech: The facts*. National Center for Women & Technology (NCWIT), 2016.

[2] David Botta, Rodrigo Werlinger, André Gagné, Konstantin Beznosov, Lee Iverson, Sidney Fels, and Brian D. Fisher. Towards understanding IT security professionals and their tools. In *Symp. on Usable Privacy and Security (SOUPS)*. ACM, 2007.

[3] Blake Burkhart. Subreddit gender ratios. http://bburky.com/subredditgenderratios/, 2017.

[4] Joseph L Fleiss, Bruce Levin, and Myunghee Cho Paik. *Statistical methods for rates and proportions*. John Wiley & Sons, 2003.

[5] International Organization for Standardization. *ISO-IEC 25010: 2011 Systems and Software Engineering-Systems and Software Quality Requirements and Evaluation (SQuaRE)-System and Software Quality Models*. ISO, 2011.

[6] André Gagné, Kasia Muldner, and Konstantin Beznosov. Identifying differences between security and other it professionals: a qualitative analysis. *HAISA*, 8:69–80, 2008.

[7] Robert M Gonyea. Self-reported data in institutional research: Review and recommendations. *New directions for institutional research*, 2005(127):73–89, 2005.

[8] Carol Grbich. *Qualitative data analysis: An introduction*. Sage, 2012.

[9] Eben M Haber and John Bailey. Design guidelines for system administration tools developed through ethnographic field studies. In *Proceedings of the 2007 symposium on Computer human interaction for the management of information technology*, page 1. ACM, 2007.

[10] Jeevitha Mahendiran, Kirstie A Hawkey, and Nur Zincir-Heywood. Exploring the need for visualizations in system administration tools. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, pages 1429–1434. ACM, 2014.

[11] Sandra R Murillo and J Alfredo Sánchez. Empowering interfaces for system administrators: Keeping the command line in mind when designing GUIs. In *Proceedings of the XV International Conference on Human Computer Interaction*, page 47. ACM, 2014.

[12] Jakob Nielsen and Robert L Mack, editors. *Usability Inspection Methods*. John Wiley & Sons, Inc., New York, NY, USA, 1994.

[13] Johnny Saldaña. *The Coding Manual for Qualitative Researchers*. SAGE Publications, 2012.

[14] William Sattelberg. The demographics of reddit: Who uses the site? https://www.techjunkie.com/demographics-reddit/, 2018.

[15] Hadi Shiravi, Ali Shiravi, and Ali A Ghorbani. A survey of visualization systems for network security. *IEEE Transactions on visualization and computer graphics*, 18(8):1313–1329, 2012.

[16] Svensk Författningssamling (SFS). *Lag (2003:460) om etikprövning av forskning som avser människor [The Act concerning the Ethical Review of Research Involving Humans]*. Utbildningsdepartementet, Stockholm, Sweden, 2003.

[17] Swedish Research Council (VR). Conducting ethical research. https://www.vr.se/utlysningar-och-beslut/villkor-for-bidrag/att-forska-etiskt.html, 2018. Accessed: 2019-02-26.

[18] Leila Takayama and Eser Kandogan. Trust as an underlying factor of system administrator interface choice. In *CHI'06 extended abstracts on Human factors in computing systems*, pages 1391–1396. ACM, 2006.

[19] Ramona Su Thompson, Esa M Rantanen, William Yurcik, and Brian P Bailey. Command line or pretty lines?: comparing textual and visual interfaces for intrusion detection. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, page 1205. ACM, 2007.

[20] Nicole F Velasquez, Suzanne P Weisband, and Alexandra Durcikova. Designing tools for system administrators: An empirical test of the integrated user satisfaction model. In *LISA*, pages 1–8, 2008.

[21] Artem Voronkov, Leonardo Horn Iwaya, Leonardo A Martucci, and Stefan Lindskog. Systematic literature review on usability of firewall configuration. *ACM Computing Surveys (CSUR)*, 50(6):87, 2018.

[22] Tianyin Xu, Vineet Pandey, and Scott Klemmer. An HCI view of configuration problems. *arXiv preprint arXiv:1601.01747*, 2016.

# Appendix

## A Survey Questions

**Page 1**

1. How much time per week (on average) do you spend directly interacting with (managing) firewalls?
   - ○ Less than 1 hour/week
   - ○ 1–4 hours/week
   - ○ 5–8 hours/week
   - ○ 9–12 hours/week
   - ○ More than 12 hours/week
   - ○ I do not directly manage firewalls
2. Can you enumerate all the firewall-related tasks that you have dealt with?
   - ○ Adding/removing firewall rules
   - ○ Examining firewall policies to understand their functionalities
   - ○ Inspecting firewall rules/policies to find errors or inconsistencies
   - ○ Other (please specify)
3. What is the PRIMARY firewall interface that you use at work?[5]
   - ○ Command Line Interface (CLI)
   - ○ Graphical User Interface (GUI)
   - ○ Application Programming Interface (API)
   - ○ Other (please specify)
4. What is your PREFERRED firewall interface?[6]
   - ○ Command Line Interface (CLI)
   - ○ Graphical User Interface (GUI)
   - ○ Application Programming Interface (API)
   - ○ Other (please specify)

**if** *answer(Q3)* = *answer(Q4)* **then**
  **go to** Page 2
**else if** *answer(Q3)* = 2 **then**
  **go to** Page 3
**else if** *answer(Q4)* = 2 **then**
  **go to** Page 4
**else**
  **go to** Page 5

**Page 2**

5. Are there certain tasks that the *%preferred%* allows you to do, which are more difficult to do using other firewall interfaces?
6. What are the strengths of the *%preferred%*, if any?
7. Can you think of any problems associated with the *%preferred%*?

**if** *answer(Q3)* = 2 **then**
  **go to** Page 10
**else**
  **go to** Page 5

**Page 3**

8. Why do you prefer the *%preferred%* to the *%primary%* when managing firewalls?
9. What are the strengths of the *%preferred%*, if any?
10. Do you see any strengths in the *%primary%*?
11. What problems do you see with the *%primary%*?

**go to** Page 10

[5] *%primary%* returns the selected option in Question 3
[6] *%preferred%* returns the selected option in Question 4

**Page 4**

12. Why do you prefer the *%preferred%* to the *%primary%* when managing firewalls?
13. What are the strengths of the *%preferred%*, if any?
14. Can you think of any problems associated with the *%preferred%*?
15. Do you see any strengths in the *%primary%*?

**go to** Page 10

**Page 5**

16. Have you ever used a graphical user interface (GUI) to manage a firewall?
    - ○ Yes
    - ○ No

**if** *answer(Q16)* = 2 **then**
    **go to** Page 6
**else**
    **go to** Page 7

**Page 6**

17. Can you name the reasons for not trying a firewall graphical user interface (GUI)?

**go to** Page 10

**Page 7**

18. Are you currently using a GUI to manage your firewall?
    - ○ Yes
    - ○ No

**if** *answer(Q18)* = 2 **then**
    **go to** Page 8
**else**
    **go to** Page 9

**Page 8**

19. Can you name the reasons for not using a firewall with a GUI and whether you see problems with GUIs?

**go to** Page 10

**Page 9**

20. For which tasks do you use the firewall graphical user interface (GUI)?

**Page 10**

21. How long have you been working as a system/network administrator?
    - ○ Less than a year
    - ○ 1–3 years
    - ○ 4–6 years
    - ○ 7–9 years
    - ○ 10 years and more
22. How would you describe your proficiency with firewalls?
    - ○ Basic knowledge
    - ○ Intermediate
    - ○ Advanced
    - ○ Expert
23. How old are you?
    - ○ 18–24 years old
    - ○ 25–34 years old
    - ○ 35–44 years old
    - ○ 45–54 years old
    - ○ 55–64 years old
    - ○ 65 years or older
    - ○ Prefer not to answer
24. What is your gender?
    - ○ Female
    - ○ Male
    - ○ Other
    - ○ Prefer not to answer

# Keepers of the Machines: Examining How System Administrators Manage Software Updates

Frank Li
University of California, Berkeley
*frankli@cs.berkeley.edu*

Lisa Rogers
University of Maryland
*lmrogers@umd.edu*

Arunesh Mathur
Princeton University
*amathur@cs.princeton.edu*

Nathan Malkin
University of California, Berkeley
*nmalkin@cs.berkeley.edu*

Marshini Chetty
Princeton University
*marshini@princeton.edu*

## ABSTRACT

Keeping machines updated is crucial for maintaining system security. While recent studies have investigated the software updating practices of end users, system administrators have received less attention. Yet, system administrators manage numerous machines for their organizations, and security lapses at these hosts can lead to damaging attacks. To improve security at scale, we therefore also need to understand how this specific population behaves and how to help administrators keep machines up-to-date.

In this paper, we study how system administrators manage software updates. We surveyed 102 administrators and interviewed 17 in-depth to understand their processes and how their methods impact updating effectiveness. We find that system administrators proceed through software updates through five main stages that, while similar to those of end users, involve significantly different considerations and actions performed, highlighting the value of focusing specifically on the administrator population. By gathering evidence on how administrators conduct updates, we identify challenges that they encountered and limitations of existing procedures at all stages of the updating process. We observe issues with comprehensively acquiring meaningful information about available updates, effectively testing and deploying updates in a timely manner, recovering from update-induced problems, and interacting with organizational and management influences. Moving forward, we propose directions for future research and community actions that may help system administrators perform updates more effectively.

## 1. INTRODUCTION

System administrators serve as "keepers of the machines," entrusted by organizations to oversee their computers, many of which are vital to an organization's operations. Their duties include regularly applying software updates in a timely manner to ensure organizational safety against crippling attacks. Failure to patch known vulnerabilities can lead to devastating consequences [13] such as the colossal 2017 Equifax data breach which exposed sensitive personal data on over *140 million* individuals [38].


*USENIX Symposium on Usable Privacy and Security (SOUPS)* 2019.
August 11–13, 2019, Santa Clara, CA, USA.

While prior studies have investigated how end users deal with software updates [18, 19, 22, 30–32, 35, 40, 45, 46, 49, 50], there has been less attention on system administrators, whose technical sophistication and unique responsibilities distinguish them from end users. Industry reports and guides on administrator patching exist (e.g., Sysadmin 101 [41]), but these lack peer-review and transparent rigorous methods. Prior academic work on system administrators is often dated and focuses on aspects of administrator operations other than updating (e.g., on general tools used [11]) or specific technical (rather than user) updating aspects. Given the critical role that system administrators play in protecting an organization's machines, it behooves us to better understand how they manage updates and identify avenues for improved update processes. We therefore set out to answer two primary research questions: (1) what processes do system administrators follow for managing updates, and (2) how do administrator actions impact how effectively they perform system updates. To answer these questions, we surveyed 102 administrators and conducted semi-structured interviews with 17 of them.

Our study determined that system administrators proceed through software updates through five main stages: (1) learning about updates from information sources, (2) deciding to update based on update characteristics, (3) preparing for update installation, (4) deploying an update, and finally (5) handling post-deployment update issues that may arise. By analyzing the factors that system administrators considered and the actions that they performed, we identified challenges that they encountered and limitations of existing procedures at each stage of the updating process. We observed problems with comprehensively obtaining relevant information about available updates, effectively testing and deploying updates in a timely fashion, and recovering from update-induced errors. We also witnessed how organizational and management influence through policies and decisions can impact the administrator's ability to handle updates effectively at multiple stages, sometimes for better, sometimes for worse. In addition, we note that while high-level aspects of software update workflows for system administrators mirror those of end users [31, 46], we found that the particular factors considered and the actions taken by system administrators are significantly different across all stages of the update process. This difference highlights the value of specifically studying the administrator population.

Our evidence-based study extends the research literature on updating practices to system administrators, a unique population. In particular, our work makes two primary contributions: first, we provide empirical grounding on how administrators update multiple machines for their organizations, examining the consequences of their actions at depths beyond prior explorations [14]. This evidence includes

insights into how their actions impact how effectively they perform software updates to better secure their systems. Second, we make grounded recommendations for improving administrator update processes through better systems for managing updates, better designed updates, and a shift in organizational policies.

# 2. BACKGROUND AND RELATED WORK

In this section, we highlight existing studies related to our research and place our work in context.

## 2.1 End Users and Software Updates

Numerous works [18, 19, 22, 30–32, 35, 40, 45, 46, 49, 50] have examined end user perceptions, attitudes, and behavior towards applying software updates. Ion *et al.* [22] and Wash *et al.* [49] found that non-expert computer users failed to recognize the security benefits of updates and frequently avoided installing them. Other studies measured the time users took to apply updates and discovered that reaching half of all vulnerable desktop [35] and mobile applications [40] took nearly 45 days and 1 week, respectively. One set of studies has examined why users avoid or fail to install software updates, discovering a variety of factors related to costs, necessity, and risks [32]. Example factors include that updates cause unexpected changes to user interfaces [19, 31, 45, 46], that updates take a long time to install [31, 46], that updates raise privacy concerns [18], and that updates cause unnecessary restarts of applications [19, 31, 45, 46].

Given that automatic updates are more effective in keeping end user systems updated than manual updates [16, 20, 35], another set of studies has examined user attitudes towards and experiences with automatic updates [30, 50]. Rader and Wash [50] identified that automatic updates with only partial user involvement (e.g., during restarts) often led to poor mental models of updating and consequently resulted in less secure systems. More recently, Mathur and Chetty [30] found that negative experiences with automatic updating resulted in users disabling auto-updates on Android devices.

While these studies have shed light on how end users deal with software updates, their findings do not necessarily generalize to system administrators, who are more technically sophisticated and operate with expanded responsibilities.

## 2.2 Administrators and Software Updates

Several studies [12, 23–25, 47, 48] have examined the workflows and needs of administrators to enable better security practices but did not focus on software updating processes specifically. Kraemer and Carayon [24] conducted interviews with 16 network administrators and security workers, identifying that organizational structures and policies played an important role in how they handled security. Kandogan *et al.* [23] discussed various stories from IT administrators about their experiences. Krombholz *et al.* [25] investigated usability problems encountered by website administrators trying to securely deploy HTTPS. Chiasson *et al.* [12] devised usability and interface design principles to help system administrators better diagnose security issues. Velasquez and Weisband [47] conducted interviews with administrators and designed a model to understand their beliefs and attitudes. In this model, the authors identified that both informational factors (e.g., quality) and system factors (e.g., ease of use) informed these beliefs and attitudes. In a follow-up study [48], the same authors found that administrators largely acquired their knowledge through practice rather than education and certification. They recommended that software developers should design tools with administrator technical sophistication in mind.

Closely related to our own work is the preliminary study conducted over a decade ago by Crameri *et al.* [14]. Although not the primary focus of their work, these researchers conducted brief surveys of 50 system administrators to learn about their updating practices. They found that nearly 70% of administrators refrained from installing software updates and that administrators tested updates on a smaller set of machines before patching their production systems. The study investigated certain aspects of administrator behavior to inform the design of their update testing system, but did not perform a comprehensive and rigorous exploration of update management. More recently, Dietrich *et al.* [15] looked at how system administrator operations could result in security misconfigurations, finding that missing and delaying software updates are among the most commonly reported security misconfigurations.

Unlike these previous studies, our work provides an in-depth investigation of system administrator practices for updating the machines they manage. Using a combination of surveys and interviews, we examine a larger sample of administrators than Crameri *et al.* [14] and provide more recent and in-depth insights into their complete update management process.

# 3. METHOD

To investigate how system administrators manage updates at scale, we conducted a qualitative study of current administrators responsible for managing updates in their organizations. Our study proceeded in two phases. In phase one, we administered a large-scale survey of administrator updating practices, whose design was informed by pilot interviews. In phase two, we conducted semi-structured interviews with administrators. We specifically sought participants who had been working at an organization with five employees or more for a period of at least one year, to ensure they had job familiarity. We restricted participation to those over 18 years old residing in the United States (US). Both study phases received approval from the Institutional Review Boards (IRBs) of our universities. Our survey and interview questions are listed in the Appendix.

## 3.1 Preliminary Phase: Pilot Interviews

In Fall 2015-Spring 2016, to inform the design of our large-scale study, we recruited seven system administrators to participate in semi-structured pilot interviews about software updates. The interview questions were developed based on prior studies on software updating [31, 46] and previous knowledge about the software update development and management process (see Appendix A for details). We recruited participants via institutional mailing lists and social media, filtering for those who explicitly dealt with software updates. All interviews were conducted over the phone via Skype and recorded. The interviews lasted between 30-50 minutes. Participants were also asked to fill out a background survey that contained general questions about demographics, the type of software or programming languages used, the types of updates they handled, and any positive and negative aspects of their job responsibilities. Participants were compensated with $20 gift cards and a chance to win a hard drive.

**Demographics:** All seven participants were male and lived in the US. They were predominantly 20–40 years of age and only one participant did not have a bachelor's degree. The majority of participants had 1–10 years of work experience as a system administrator.

**Analysis:** We transcribed all pilot interviews and three coders used inductive thematic analysis [42] to derive the following over-arching themes in administrator update management: finding information about available software updates, testing and preparing for updates, deploying updates, and monitoring for update-triggered issues post-deployment. We used these themes to design questions for our study's two phases.

## 3.2 Phase One: Survey

Based on the pilot interviews, we constructed a survey asking about a participant's organization and responsibilities (e.g., size of organization, number of machines managed), how they manage the security of their systems, how they handle each stage of the update management process, and what works well and poorly for them (see Appendix B for details). The survey consisted of 41 questions and took approximately 15 minutes to complete. We recruited system administrators in September and October of 2017 using social media, blogs associated with our research labs, and Reddit [7]. In addition, we recruited administrators attending the 2017 Large Installation System Administration Conference (LISA) by distributing fliers about our survey and providing a computer at the venue where administrators could complete the survey. As an incentive, we entered administrators who participated into a drawing for a Samsung S8 phone. In total, 102 system administrators completed the entire survey. We note that we recruited 22/102 survey participants at the LISA conference and the rest from online.

**Data Analysis Method:** The survey consisted of multiple-choice and open-ended questions. We focused our analysis on questions pertaining to software updating, as our survey also contained several less relevant questions on other security practices. We analyzed open-ended questions using open coding, identifying themes in the question responses [51]. Two researchers independently developed a set of codes across all questions and met to converge on a final codebook. Then, each researcher independently coded all question responses using that codebook. We had 199 codes with 611 coded segments in total, discussing themes of interest such as "Testing", "Update Issues", "Addressing Update Issues", "What Works Well", and "What is Challenging". We use Kupper-Hafner inter-rater agreement scores [26] to quantify the consistency of the coding, finding an average agreement of 0.83, indicative of largely consistent coding. The survey coders met and converged upon the final codes for all open-ended question responses.

## 3.3 Phase Two: Semi-Structured Interviews

Using the themes identified by our pilot interviews, we developed a guide for conducting semi-structured interviews with system administrators. The guide contained questions about a participant's demographics and job, and their update management process (see Appendix C for details). Throughout Fall 2017, we recruited 17 interview subjects through the same channels as with the survey. All but one of our subjects participated in the survey as well. Interviews ranged from 1 to 3 hours long, were conducted in person or over Skype, and were recorded. We compensated participants with a $20 Amazon gift card.

**Data Analysis Method:** Using transcriptions of the recorded interviews, we developed a codebook for the responses through regular peer review meetings, based on the themes of interest for the interviews such as "Job Responsibilities", "Update Importance", and the various update stages, including "Seeking Update Information", "Deployment", "Testing", and "Update Issues". The codes were initially created by one team member and refined by group discussions and consensus [51]. Two coders independently coded the interview responses using the resulting codebook using inductive thematic analysis [42]. We had 347 codes with 1447 coded segments in total. Calculating inter-rater reliability for such qualitative coding of non-survey data has been shown to be difficult because of the nature of assigning multiple codes to data and inherent biases of coders [10]. For completeness, however, we randomly sampled 6/17 transcripts and computed an average agreement percentage between the two independent coders of 0.77, indicating high consistency. We discussed points of disagreement and ensured that the resulting themes discussed in the paper were in line with both team members' interpretations of the data.

## 3.4 Participant Demographics

Here we present the demographics of the 102 survey respondents and 17 interview subjects.

### *3.4.1 Respondent Characteristics*

The population was male-dominated; only 6/102 survey and 2/17 interview subjects were female. The most common age bracket was 26-35 years old, containing 43/102 survey participants and 8/17 interview subjects. Other common age brackets were 36-45 years old (24 survey and 4 interview participants), and 46-55 years old (14 survey and 2 interview subjects). Most administrators had some higher education; 57/102 survey and 10/17 interview participants had a bachelor's degree while 37 survey and 5 interview participants had some college education but no degree. Salaries varied widely, evenly distributed primarily between $35,000 to $150,000 (accounting for 93/102 survey and 14/17 interview participants). Survey respondents had a median of 11 years of experience, ranging from 1 to 35 years. In contrast, interview subjects had a lower median experience of 6 years, although the range was similar (1-34 years).

### *3.4.2 Organization Characteristics*

About half of our study participants (56/102 in the surveys and 8/17 from the interviews) worked at larger organizations with over 500 employees. In comparison, only 13/102 survey and 2/17 interview participants worked for small organizations with fewer than 50 employees. In total, 22 survey respondents did not indicate the number of hosts they managed (all interview subjects did provide a response). However, the remaining typically oversaw many machines: only 12/102 survey and 3/17 interview participants maintained fewer than 100 hosts, while 36 survey and 8 interview subjects indicated they administered between 100-499 machines and 22 survey and 5 interview participants said they handled over 1000 machines. Servers were the most common type of machine managed, handled by 96/102 survey respondents. Over half of the administrators also dealt with desktops (63), routers (60), and laptops (57). Our participants maintained primarily Linux (73) and Windows (71) machines, and less so Macs (44).

## 3.5 Limitations

Studying system administrators is challenging as they are a specialized population that is difficult to recruit compared to end users. Thus, our study's approach may have limitations.

1. As administrators are often paid well, our study's participation compensation may not have influenced their decisions to contribute. Instead, those more ideologically motivated may have donated their time.
2. Due to our recruitment method, our study participants may not be representative of system administrators in general. For example, we only studied individuals from the US, so our findings may not apply globally. Similarly, we only recruited administrators fully employed by an organization, which does not capture those working part-time or as contractors.
3. Our results reflect our study's sample, which skewed towards certain demographics (e.g., males). Similarly, we recruited many of our participants via Reddit and the LISA conference. These subpopulations may exhibit certain skewed characteristics. For example, those attending the LISA conference may operate with a larger budget (covering conference expenses).
4. Our surveys and interviews contained open-ended questions.

During our analysis, we provide the number of study subjects who gave a particular response to these open-ended questions (and indicate when results are obtained from such questions). However, we caution that such counts are not necessarily reliable indicators of real-world prevalence. In particular, we cannot assume a respondent does not act a certain way just because they do not mention such behavior, as they may have simply focused on alternative discussion topics.

5. Our study is an exploratory one that focuses on the processes system administrators use to manage software updates. However, we did not investigate all updating aspects in depth. For example, we did not explicitly solicit recommendations from our study participants on how to improve updating tools and methods, nor did we tease apart the differences in updating between different types of organizations or machines. Moving forward, our study can help inform the design of broader quantitative explorations of these updating dimensions at scale.

# 4. OVERVIEW OF FINDINGS

From the responses to our system administrator surveys and interviews, we determined that administrator software update workflows consisted of five primary stages. These five stages, as illustrated in Figure 1, are: (1) learning about updates from information sources, (2) deciding to update based on update characteristics, (3) preparing for update installation, (4) deploying the update, and finally (5) handling post-deployment update issues that may arise. For each stage, our analysis determined the factors that system administrators considered and the actions that they conducted (also listed in Figure 1). This data affords us insights into the challenges that administrators encountered when updating and limitations of existing procedures. In Section 11, we discuss recommendations for improving administrator update processes grounded in our findings. We also compare how update workflows differ for system administrators versus end users in Section 11.1, identifying significant differences.

In the following sections on update stages, we explore how system administrators proceed through each stage and the security implications of their behaviors. Throughout the results, we designate quotes from survey respondents with *S* and interview participants with *P*.

# 5. STAGE 1: LEARNING ABOUT UPDATES

In both our surveys and interviews, participants reported that—before deploying software updates—they first had to learn about available updates and then make decisions about which updates to handle. We note that while automatically initiated updates circumvent the need to find and digest information, many of our study participants did not find them universally suitable. Thus, for our participants, it was still important to process update information efficiently.

## 5.1 Update Processes

We asked our study participants about how they discovered the updates they applied. In our survey, we asked a closed-ended question with 11 possible options and a free-form response (as shown in Table 1), while our interview question was open-ended. In total, 99/102 survey participants and all 17 interview subjects responded. The types of information sources discussed by interview subjects overlapped with our survey question options, but we note that the distributions among survey and interview participants differed, likely due to the open-ended nature of the interview question.

As shown in Table 1, our participants relied on various types of information sources. Most survey respondents reported a median of 5.0 different types of sources, and a quarter reported using seven or more types. (We do not report the same counts for interview data given that open-ended responses are not necessarily comprehensive nor indicative of prevalence, as discussed in Section 3.5.) This large quantity of source types suggests that update information is highly dispersed, requiring administrators to diligently peruse a variety of outlets to stay informed on available updates. Some interview participants described sourcing information in this manner as non-ideal, as typified by P5's discussion on discovering updates that patched newly identified vulnerabilities: *"There's not always a canonical*

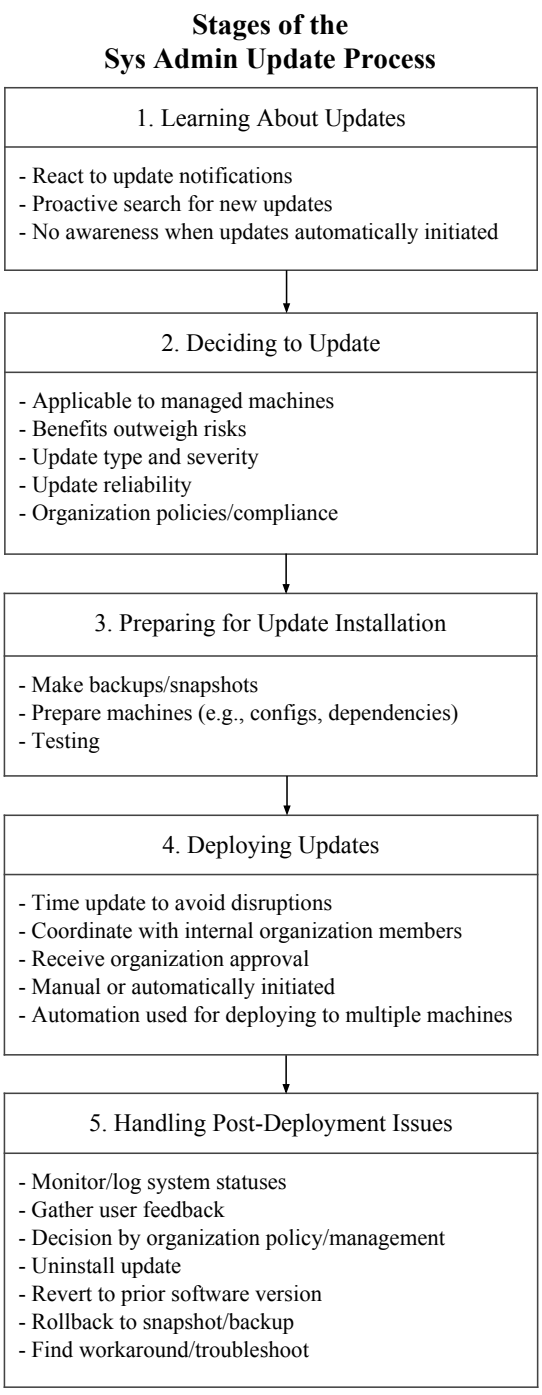



**Figure 1: Our study identified five primary stages of the update process for system administrators. We list the salient considerations for each stage.**

**Table 1: Sources used for discovering available updates.**

| | Source for Update Availability | # Survey Responses | # Interview Responses |
|---|---|---|---|
| 1. | Security advisories | 80 (78%) | 4 (24%) |
| 2. | Direct vendor notifications | 72 (71%) | 11 (65%) |
| 3. | Professional mailing lists | 54 (53%) | 7 (41%) |
| 4. | Online forums | 53 (52%) | 7 (41%) |
| 5. | Alerts from software | 41 (40%) | 10 (59%) |
| 6. | News | 40 (39%) | 5 (29%) |
| 7. | Blogs | 39 (38%) | 5 (29%) |
| 8. | Third-party services | 28 (28%) | 0 ( 0%) |
| 9. | RSS feeds | 22 (22%) | 3 (18%) |
| 10. | Project mailing lists | 21 (21%) | 0 ( 0%) |
| 11. | Social media | 18 (18%) | 1 ( 6%) |
| 12. | Other | 9 ( 9%) | 3 (18%) |
| 13. | No Answer | 3 ( 3%) | 0 ( 0%) |

*place to go for a web advisory. When these vulnerabilities get found on the Internet, they might affect you, it could be announced on the Apache web server mailing list, it could be on the Ubuntu server list, it could be a topic on Server Fault. There's a lot of places."* Also, not all sources were ideal. For example, P13 stated that *"sometimes if there's a really critical vulnerability, email's not the most real-time method of getting things going."*

## 5.2 Impact on Updating Effectiveness

Our study participants revealed that they each relied on a diverse set of methods for retrieving update information from multiple sources. Due to the lack of a centralized source of information, we note that it is possible that some system administrators may lack the full coverage of relevant information if they miss an important source. We also observed that administrators used some sources that require active retrieval and digestion, such as news articles, blog posts, forums, and social media. These sources may require more time and effort, compared to sources that push information directly to the administrators, such as direct vendor notifications or mailing lists. Our study ultimately does not concretely reveal how comprehensive or effective administrators are at update information retrieval, but suggests that this is a nontrivial task for many.

# 6. STAGE 2: DECIDING TO UPDATE

For the second stage of their updating process, administrators in our study filtered update information to decide if they should deploy an update. This was a nontrivial task because of the profusion of update information from a variety of sources.

## 6.1 Update Processes

In our survey, we asked respondents about which types of updates they most frequently apply. In our interviews, we asked our participants how they determined which updates to deploy, and which types of updates they considered important. From the responses, we observed five primary factors that our participants discussed for assessing the cost-benefit trade-off of applying available updates. Our interview question was open-ended, so this set of factors may not be comprehensive or indicative of prevalence, as discussed in Section 3.5.

**1. Update Type:** In a closed-ended question, we asked our survey participants which updates they regularly installed: security or non-security related updates. In total, 97/102 administrators regularly installed security updates, whereas only 63/102 administrators did likewise for non-security related updates. (3 respondents did not answer.) We similarly asked our interview subjects an open-ended question about their views on which updates were important or not. Most interview participants (15/17) said that they considered security updates to be vital, but they disagreed on the importance of other updates; 7 administrators considered them important, whereas 5 administrators did not, often feeling they could be disruptive. For example, in a quote that is typical of what we heard, P16 explained: *"Least important, anything that's like feature updates or considered upgrades. I don't really want new features, because new features mean new problems, so I just want to get the security stuff tucked away."* Thus, our study participants typically found security updates important to apply.

**2. Update Severity:** In an open-ended interview question on how administrators decided to apply an update, the severity of the issues addressed by an update was a factor discussed by 9/17 interview participants. In a canonical example, P13 prioritized updates to *"Only critical security ones...It mostly depends on the severity and what the risk is."*

**3. Update Relevance:** When discussing their process for deciding to apply an update, five interview participants (29%) explicitly described update information overload, where much of the information they acquired did not apply to their machines. As a result, they said that they had to tediously filter out unnecessary information (or possibly avoid overly verbose feeds altogether). For example, P6 thought that *"Sometimes there's an overabundance of information...there are some products, things like that, that we don't use here. So I have to actively filter that out myself."* Others described receiving multiple emails about specific upgrades (e.g., Linux patches simultaneously released in batches) and how these emails were easily lost or hard to process in an overflowing inbox.

**4. Update Reliability:** Three interview subjects brought up known update issues as another factor in determining whether to update. For example, P11 cared about the update quality, saying *"a reliability score of an update would be my number one [update characteristic]."*

**5. Organizational Factors:** In many cases, organizational or management policies and decisions influenced or even dictated the update decision. We discuss in more detail in Section 10.

## 6.2 Impact on Updating Effectiveness

We found that system administrators prioritized updates that fixed security (or other severe) bugs. However, many software updates bundle bug fixes with feature or performance changes, including popular software such as the Mozilla Firefox Browser [4] and the Apache HTTP web server [2]. This entanglement suggests that it is challenging for administrators to specifically address the most urgent software problems without contending with other potential changes. Additionally, certain update characteristics (e.g., update reliability) were important to our study participants in deciding whether to apply an update. However, updates may not contain information to assess such characteristics (e.g., Firefox [4], Apache HTTP daemon [2]), or provide too much irrelevant information (described by study subjects as information overload), making it challenging for administrators to make informed updating decisions.

# 7. STAGE 3: PREPARING FOR UPDATE INSTALLATION

After identifying appropriate updates, our study participants reported that they had to make preparations for installation, which fell into three over-arching categories. First, administrators frequently *made backups/snapshots* in case problems arose through the updating process. Second, they *prepared machines* when necessary, such as by changing configurations or dependencies. These actions were

often necessary due to the manual nature of many updates. Finally, they often extensively *tested* updates for unintended side-effects or bugs. Here, we focus on the testing considerations of administrators as we cover the other two considerations in the remaining sections.

**Threat of Bad Updates:** We asked our survey and interview participants to describe their experiences with problems caused by updates on the machines they managed. In a closed-ended survey question, we asked how frequently an administrator encountered a problematic update. Of the 98/102 survey respondents that answered, all but 2 said that they had encountered bad updates; 54 indicated this happened infrequently, 36 found problems every few update cycles, and 6 said most update cycles produced complications. When asked an open-ended question on whether they tested updates and why, our interview subjects expressed the same sentiments on update risk; 8/17 recounted running into a recent faulty update. While the participants' recollections may not have been entirely accurate, it reflected a general sentiment among them that updating comes with non-trivial risks that they should manage. In the worst case, the negative experiences drove administrators towards fewer updates: *"I stopped applying updates because it was becoming more of a problem to apply them than not to. Production machines, they don't get updates"* (P12). Such behavior can leave hosts riddled with security vulnerabilities and ripe for compromise. To combat the risks of bad updates, many of our study participants engaged in the time-consuming process of update testing.

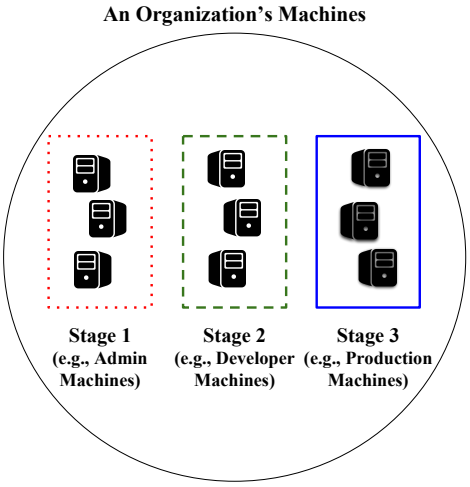



**Figure 2: Staggered Deployment Testing: The system administrator allocates machines into stage groups, and updates stage by stage, waiting between each stage for update issues to manifest. If they arise, the administrator halts deployment and investigates the issues. For example, an administrator at a software company might first group only machines that they use as the first stage, then group developer machines as a second stage, and form a final stage of production machines.**

## 7.1 Update Processes

Both our surveys and interviews contained an open-ended question asking respondents about what testing they do for updates, if any, and why. The majority of our participants (83/102 survey respondents and all 17 interview subjects) indicated they tested updates. (Seven survey participants did not respond.) Among those who tested, 22 survey participants and 3 interview subjects discussed only ad-hoc testing methods (e.g., testing basic software functionality) without discussing any strategies in detail. For the remaining administrators, we found that testing strategies varied but fell into two general classes: *staggered deployments* and *dedicated testing setups*. Regardless of the chosen strategy, testing was often a pain point for administrators: in open-ended survey questions on what works well and poorly in an administrator's updating process, only 14/102 survey respondents recounted positive testing experiences, and 12 reported that developing a reliable testing workflow was the most difficult aspect of updating. Thus, many of our study participants found it challenging to develop a dependable testing process.

**1. Staggered deployments.** When staggering update deployment (as illustrated in Figure 2), administrators in our study described separating their machines into multiple groups, deploying updates to a group at a time, and waiting some time between each stage of deployment to observe if update issues arise. In an example that summarizes this approach, S72 said that they first *"install on non-important machines and let them bake for 1+ months."* This strategy, which merges update testing and deployment, was the most commonly used among our study participants, leveraged by 43/102 of the survey respondents and 11/17 of the interview subjects.

We identified three different ways that participants used to grouped machines in each stage. First, 22 survey respondents and 4 interview subjects categorized machines into priority levels, testing updates first on lower priority machines. A second approach (10 survey respondents and 2 interview subjects) was to test first on the machines of end users who opted into assisting with update testing. For example, P10 talked about deploying updates to volunteers for a week prior to company-wide rollout, a strategy many spoke of using because: *"They're very good at reporting things that have gone wrong."* A final less-frequently used strategy was to pick pilot groups at random, only discussed by one survey participant and two interview subjects. While P11 selected machines completely at random, independent of the user, P5 chose randomly with more nuance: *"Usually, it's randomly picking something that I know is active but not the most active machine out there. If I pick something that nobody's using for anything, then, that's not a good place to test it. But, it's also not one of our highest risk servers."*

Our survey participants typically did not indicate how they monitored for update problems during staggered deployment, although four respondents mentioned gathering user feedback from those who piloted updates. Interview participants told us that they monitored how well updates were applied through monitoring software (6/17), lack of error messages (6/17), checking the machines for compliance (2/17), and user feedback (1/17).

**2. Dedicated testing environments.** Our survey participants often mentioned a dedicated testing setup, where they used machines provisioned specifically for testing (30/102 survey respondents) or relied on a testing or quality assurance (QA) team (9/102). (Five survey respondents used both approaches.) Figure 3 illustrates this process. Among interview participants, 8/17 used dedicated test servers, with two also having a QA team. S29 captured the gist of this approach: *"We test in a lab/test environment that has similar functions as our production environment. We do this to ensure we get accurate and reliable results that won't break our end users' applications."* Similarly, S19 gave an example of how QA teams conducted testing: *"For some third-party software (issue tracking, artifact management, etc.), our QA department has scripts to exercise business-critical functionality."* We note that 16 survey respondents and 5 interview subjects with dedicated testing also used staggered deployment, suggesting that often participants felt that dedicated testing was not sufficient by itself.

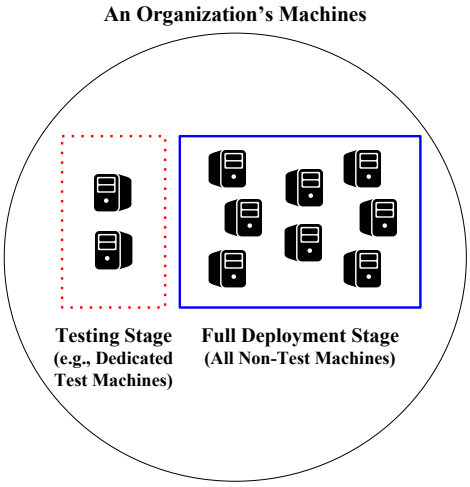



**Figure 3: Dedicated Testing Environments: The system administrator evaluates an update in an environment configured specifically for testing (e.g., test servers). If they do not discover update issues, they fully deploy the update (potentially via staggered deployment).**

Most of our study participants did not elaborate on the specific evaluation done in dedicated testing setups, although some mentioned automated software testing and manual investigation to confirm that critical software functionality remained. We note that no more than three survey or interview participants explicitly mentioned any particular method though, so further exploration on dedicated testing details is warranted.

**4. No testing at all.** A minority of survey respondents (10/102) did not test updates at all, and an additional two respondents indicated that they skipped testing on some of their systems as it was infeasible, without discussing testing on other systems. No interview subjects avoided testing. Three of the survey respondents who skipped testing did not provide their reasons. However, two survey respondents indicated they lacked the time, and three others deemed updates in their environment to be low-risk enough to deploy without testing. For example, S43 acknowledged, *"It is a poor habit but I don't ever experience any issues with Microsoft updates, so I see no reason to wait before applying them."* In another instance, a survey respondent skipped updates because testing on a diverse set of hosts seemed impractical, stating that with *"Too many different environments, would need to test a dozen different ways before deployment"* (S34). The final test-less respondent S37 stated that they skipped testing because *"security patches are a requirement, if it breaks something it gets fixed downstream."*

## 7.2 Impact on Updating Effectiveness

Those participants who used the staggered deployment testing strategy avoided the need for dedicated testing resources (although some used both strategies). However, we note that an important downside of staggered deployment for participants was that it could significantly delay updates to hosts in later stages. Some study participants indicated this delay could be on the order of weeks or months. Notably, administrators often spoke of updating production machines last, which is particularly concerning as these servers often directly interacted with external entities and hence, potential attackers.

Some of our administrators preferred dedicated testing environments for evaluating updates in a low-risk setting. However, we note this strategy requires additional computing resources or employees specifically for testing. In addition, we heard from participants about the challenges in replicating nuances of real-world deployments in testing environments. Ultimately, update testing was a challenging endeavor for most of the administrators in our study, driving some to even bypass testing.

# 8. STAGE 4: DEPLOYING UPDATES

Our study participants had to develop methods for deploying updates across the many machines under their purview.

## 8.1 Update Processes

Specifically, our study participants had to determine how to deploy updates and when to do it.

**1. How to Deploy?** In a closed-ended survey question, we asked survey participants whether they deployed updates manually, wrote their own programs or scripts to deploy updates, used third-party update management software, enabled automatic updates, or deployed in an alternate fashion (with a free-form response). Based on the 99/102 survey respondents who answered, we observed that administrators often lacked a single unified system for deploying updates. While 34 survey participants used a single method, the rest used multiple, with a median of 2 methods. We asked interview subjects an open-ended question on how they install updates, and interview participants also reported a mixture of deployment methods.

A majority of survey respondents used third-party update managers[1] (64/102), as did 12/17 of the interview subjects. P14 described their use of the update management software Ansible [1], explaining *"with Ansible you would just specify a list or subsection of a list of machines to run a particular command or update and it would run all of those in parallel on each of the machines and return the status of the request."* Some interview participants felt these tools could be improved to take snapshots of their systems and better indicate missing updates for specific machines.

Almost half of our participants (50/102 survey respondents and 7/17 interview subjects) created custom scripts or programs to automate the deployment process, while 44/102 survey participants and 2/17 interviewees enabled automatic updates for some software packages. Manual updates were still frequent though, conducted by 40/102 survey respondents and 4/17 interview subjects. One consequence of the heavy use of scripting and manual actions was the issue of legacy systems and processes. For example, P7 illustrated one scenario, saying *"If there's a legacy system in place and Jeff the sysadmin is the only dude who even knows how to run the scripts for that, or whatever service is running on there, you know, God forbid Jeff gets hit by a bus."*

**On Automation:** In response to open-ended survey questions on what aspects of an administrator's update management process work well and which are challenging, many study participants spoke of the importance of automation in the update deployment process. In a representative quote, S62 explained: *"Automating the process is essential for any environment with more than 10 endpoints as it greatly reduced the time involved and also improves the frequency of patch application."* S19 agreed in their response to the same survey question, stating *"There is no way our small team could manage this many machines without [automation]."* However, implementing automation often required significant effort. P15 stated they did not initially automate due to *"just the amount of time it would take*

[1] Tools mentioned included Ansible [1], SCCM [33], Chef [3], Spiceworks [8], Puppet [6], Terraform [9], and WSUS [34].

*to implement all the automation."* This participant did later deploy automation, stating it took them *"three months, to get it right."*

Even with the benefits of automation, our survey participants also highlighted that many situations still required manual actions, some in preparation for update installation (as mentioned in Section 7). For example, S14 sometimes still performed manual updates because *"Major OS updates require more manual intervention, such as updating custom scripts, updating or rewriting configuration files, or updating third-party tools."* In the interviews, some subjects mentioned that automation was not always desirable since update issues could arise unexpectedly.

Dependency and compatibility concerns posed particular problems for automation. In a prototypical example, S62 struggled with *"Maintaining compatibility with software that depends on platforms like Java/.Net/etc. Vendors tend to lag behind the platform by at least 1-2 release cycles preventing us from updating to the latest version."* Additionally, host heterogeneity (e.g., different software versions) complicated update deployment as illustrated by the following typifying example: S86 found deploying updates difficult when *"pushing to multiple versions of Linux with only one tool."*

Thus, while automatic updates and deployment automation was helpful and important for our study participants, they often could not fully automate updates across their machines due to some of the above reasons.

**2. When to Deploy?** In open-ended interview questions on how administrators deployed updates and whether they had to notify anyone about the update, our interview subjects frequently discussed the need to minimize disruptions for users and updated machines. (Our survey did not contain equivalent questions.) One strategy for mitigating disruptions (used by 13/17 interview subjects) was to update along a predictable schedule, such as P10's weekly patching program, so that users were not caught off-guard by the update timing. Another strategy mentioned by 12/17 interview participants was to update during off-hours. We also observed that organization and management decisions could dictate when updates occurred (described in Section 10).

In many cases, communication and coordination with those affected by an update were vital. This sentiment is best exhibited by P10's (who followed a weekly update schedule) discussion of their coordination efforts: *"On a given week, your machine might get software and it might reboot. We have a communication program that goes along with that, that we send out to the units about what's happening this week."* In a contrasting but similar example of coordination, P5 told us that they based update timing on user preferences: *"You send out an email to people and see what time works best for them. Usually, they can identify a time that is going to be idle for them or lower use than regular."*

## 8.2 Impact on Updating Effectiveness

Challenges in implementing automation for update deployment forced many of our participants to perform manual updates. In addition, administrators in our study often eschewed automatic updates so they could make proper preparations. We note that these manual actions could result in slower update rollouts leaving machines exposed to bugs and vulnerabilities for a longer duration. Also, manual updates may require further effort and be more prone to human error, potentially resulting in misconfigurations or functionality regressions. For our participants, the need to time updates in coordination with organization members or policies further widened the vulnerability window for machines.

# 9. STAGE 5: HANDLING UPDATE ISSUES AFTER DEPLOYMENT

Unfortunately, update testing did not always prevent issues from arising post-deployment. We asked our survey participants an open-ended question on how they became aware of problems caused by installed updates. In total, 56/102 survey participants found out about some update issues through user or client complaints, while 21 discovered problems through monitoring updated hosts. We further asked both our survey and interview participants open-ended questions about how they handled these post-deployment problems.

## 9.1 Update Processes

Of the 93/102 survey respondents that answered, only 3 indicated they lacked a process for managing post-deployments issues. From the interviews, 11/17 subjects reported recently running into post-deployment problems.

For the administrators that did deal with update complications, the most common approach was to uninstall an update. In total, 48/102 survey participants used this strategy, with 6 mentioning that they did so with custom scripts and 20 using third-party software or an update manager to do so (the rest did not specify). Similarly, 6/17 interviewees mentioned having to uninstall updates to resolve update problems. Another common approach was to revert to a previous snapshot or backup of the software or system. This strategy, used by 35/102 survey respondents and 7/17 interview subjects, did require proactive steps in preparation for update installation (namely, making a backup), as mentioned in Section 7. In an example of the forethought required of administrators, S5 discussed their backup strategy: *"I take an image of the entire disk once a month for non-critical machines and daily for critical machines."* Other rollback strategies mentioned less frequently during the surveys and interviews included downgrading to an earlier version of the software (possibly undoing several update cycles), manually negating an update's changes, or reverting to a mirrored/parallel environment.

The prior strategies all involve returning to a pre-update state, which can leave machines without patches for new vulnerabilities. Some administrators preferred to keep the updates in place, with 15/102 survey participants and 1/17 interviewees saying they attempted to find workarounds for problematic updates. Of these, 4 survey participants said they never roll back, focusing on keeping updates in place while managing any issues. Also, 7/102 of our survey participants relied on vendor assistance in resolving update issues.

## 9.2 Impact on Updating Effectiveness

After deploying an update, if problems arose, our study participants tended to revert to a functional but insecure prior state, demonstrating that they prioritized functionality over security. This behavior also suggests that system administrators found it difficult to identify workarounds or fixes for update problems, whether by themselves or via the software vendors.

# 10. ACROSS STAGES: ORGANIZATION AND MANAGEMENT INFLUENCE

A significant theme that emerged from our study participants was the important role that an organization's internal policies and management could play in update decisions. This theme provides new evidence extending the work by Dietrich *et al.* [15], who also observed that organizational factors impacted how administrators handled system misconfigurations.

We briefly note that we explored whether organizational structure, such as the number of employees or machines managed, affected our participant's update management practices, particularly related

to different testing and deployment strategies. To do so, we compared the distributions of the organization size and the number of machines managed between those adopting different updating behaviors. We used the Mann-Whitney-Wilcoxon test [29], with a p-value threshold of $\alpha = 0.05$, to determine if the distributions statistically significantly differed. However, we did not identify any significant differences; thus, the organizational structure did not appear strongly correlated with any particular update process.

## 10.1 Update Processes

Across responses to various open-ended questions, our study participants discussed situations where organizational policies and management affected updating practices.

**1. Free Reign.** In some organizations, administrators had decision-making authority and could apply updates as they saw fit. However, this put the onus on the administrator solely to keep machines secure. P11's company exemplified this approach: *"I don't have to run junk through a bunch of red tape to do anything. I just do it, knowing the consequences; things could break, could cause a lot of problems and lose a lot of money, but that's just part of having the responsibilities of that job I have. If I want to push out updates to all 1,800 machines, I don't have to really answer to anybody."*

**2. Organizational Oversight.** In other cases, administrators in our study told us they had to get management buy-in before taking certain update actions. A quote from S26 characterizes this setup, as they talked about applying updates only after management approval because *"I will be fired if I do so before I can convince management."* Similarly, in another representative example, S70 discussed that their update promptness was often delayed because *"Mostly the business being incompetent and not approving the work to go ahead. If it was up to me, [updates would be installed] as soon as they are released and after testing."* This setup often made updating challenging for participants. For example, S14 had to fight for maintaining Windows updates, as management felt that those updates were not trustworthy. These disagreements between administrators and management appeared to result in updating practices that the administrators in our study did not always support.

In some cases, organizational policies dictated the actions of the administrators. A canonical example from S37 illustrates the pressure on their update deployment timeline: *"Policy and compliance require deploying them within 5/10/30 days depending on severity."* In another example quote, P15 explained that their organization's requirements determined the priority of different updates: *"We have compliance implications around getting security updates out, so that's one. We have an organizational mandate to deliver a stable platform, so stability updates set prioritized as well."* With a potentially less secure outcome, P12's organization decided to reduce the frequency of machine updating, because *"that's just more of a decision that we've made as a business that...it's just better not to introduce a problem."*

Several study participants also commented on another important organizational decision: the budget allocated for system administrator operations. For example, S21 said they lacked the time for managing updates but *"My company won't let me buy anything to help with automatically deploying."* Similarly, P16 said that they lacked the budget for obtaining good software to handle updates until demonstrating their network's insecurities to management.

## 10.2 Impact on Updating Effectiveness

Organizational freedom allowed some of our study participants to more effectively apply updates, but placed the burden of security on their shoulders alone. We note that such freedom could result in ad-hoc decision making by administrators, potentially resulting in poor practices, or decisions that could negatively impact other aspects of an organization, such as the reliability or availability of an organization's production systems.

By contrast, requiring management approval complicated the update process for many system administrators and could delay or prevent the application of updates. Such barriers also drove down the updating frequency for those administrators who told us they can only request approval for the most severe updates, and often, some skipped less severe updates to avoid the hassle of getting approval.

# 11. DISCUSSION

Our study of system administrator software updating identified how administrators perform updates and the security implications of their behaviors. Future user studies on administrator software updating could extend our work to develop a richer model of update decision-making processes, investigate how updating differs for different types of organizations and machines, explore the effects of organizational policies on updates in more depth, and identify concrete steps for improving updating tools and interfaces. In this section, we synthesize our findings to identify how software updating differs between system administrators and end users, and how we can help administrators better keep machines updated through recommendations grounded in our results.

## 11.1 Comparison with End User Software Updating Practices

Prior work on software updating behavior has primarily studied end users. From synthesizing and comparing with the results from existing studies [19, 30–32, 45, 46], we find that end users follow similar stages of the updating process, but with differing considerations at each stage. Overall, we observe that administrators performed more sophisticated tasks (e.g., testing) and had unique aspects of their workflows as a result of managing numerous heterogeneous machines within an organizational context (e.g., staggered deployment, organizational influences). For each of our five updating stages (summarized in Figure 1), we highlight the salient differences between end user and system administrator considerations.

- **Stage 1 (Learning):** Administrators relied on a diverse set of update information sources, including those from proactive searching. In comparison, end users primarily learned about updates through notifications or alerts from within their software and rarely sought updates by themselves [31, 46].
- **Stage 2 (Deciding):** Like end users, administrators in our study considered the benefits and risks of an update [19, 31, 32, 45, 46]. However, our participants had the additional facet of determining if and which updates affected the potentially heterogeneous hosts in their organization. Some administrators also had to abide by organization policies.
- **Stage 3 (Preparing):** We observed that update-induced issues concerned both our study participants and end users [19, 30, 31, 45, 46]. As a result, end users either avoided updating, updated after making backups, or dealt with update issues only after applying [46]. In comparison, administrators took more extensive preparatory steps, including backing up and snapshotting systems, modifying software configurations and dependencies, and testing updates before applying them.
- **Stage 4 (Deploying):** As administrators in our study deployed updates at scale, unlike end users, they had to consider the interruptions and downtime on machines they served, often requiring coordination with other organization members or or-

ganization approval to take actions. They also often employed automation to scale up their updating tasks.

- **Stage 5 (Remedying):** When updates caused issues, both populations employed similar high-level remedies (e.g., uninstalling updates, finding workarounds) [46]. However, administrators in our study had to contend with the challenge of identifying update issues across numerous machines that they updated, requiring them to consider monitoring systems and feedback from these machines' users. Additionally, as these issues could affect organizational operations, organization factors influenced how administrators handled these situations.

## 11.2 Reducing the Burden of Update Information Retrieval

In Section 5, we learned that information on software updates is widely dispersed across various sources. Our findings suggest that helping administrators more easily identify relevant updates for their machines would simplify their updating efforts and increase the likelihood of prompt updating. One solution could be to standardize and consolidate update information at a centralized repository (similar to efforts on aggregating vulnerability information [36]), providing a singular destination for identifying available updates.

Another intriguing approach is through outreach campaigns that inform administrators about severe vulnerabilities and promote updating to patch the security holes. Several recent works [17, 27, 28, 43, 44] have investigated the benefits of reaching out to the administrators of machines with publicly visible security issues, finding that the notification efforts resulted in a significant improvement in the remediation of the security problems. However, they also identified hurdles in contacting all administrators and promoting corrective actions, and there remain important research questions such as how to effectively deliver messages, whom to contact, how to establish trust with recipients, and how to incentivize remediation. Thus, we recommend further research on improving administrator notifications to overcome existing challenges and identify best practices.

## 11.3 Simplifying Update Decision-Making

Our findings in Section 6 indicate that administrators prioritize updates with certain characteristics (e.g., update severity), so standardizing update information to consistently include such characterizations would aid them in their decision-making. In particular, administrators differentiate update types. Thus, there is value in splitting all-inclusive updates into updates specific to one type of patch, as also recommended by prior work on end user updates [31,46]. For example, software vendors could bundle security patches separately from feature patches. With this segregation, administrators can better prioritize the updates they apply (e.g., security fixes). However, we recognize that splitting updates could complicate software development and release. Future work could therefore explore how best to separate and enable updates of different types, from both the software developer and administrator standpoints.

## 11.4 Improving Update Deployment Processes

There remains a salient need for advancements in the update tools that system administrators rely upon, as we observed that administrators encountered various hurdles throughout the preparation and deployment of updates (Sections 7 and 8), and the handling of post-deployment problems (Section 9). For example, the notion that automatic updates would solve the patching problem is overly simplistic, as our findings demonstrate the complexities of the updating process (particularly with situations still requiring manual actions, as discussed in Section 8).

While technical developments are certainly needed, we also lack a deep understanding of the usability of these tools. Therefore, the usable security community could contribute explorations into how administrators use update tools and how their interfaces could be improved. For example, our findings (in Section 8) indicate that many administrators use third-party update managers. What information do they display before, during, and after update deployment, and what missing information (such as on dependencies or affected configurations) would streamline administrator workflows if provided?

One notable deployment issue our administrators faced was timing updates to avoid operation interruptions. We believe that dynamic software updating [21] (DSU), a method that allows for live updates without restarts or downtime, could help with side-stepping update timing concerns. While it has not yet been widely deployed, the approach is promising as some major systems have adopted it, such as with the Linux kernel extension Ksplice [5]. However, we have little understanding so far of how using DSU systems affect developers writing patches and administrators operating such systems. For example, the use of DSU systems can result in complex data representations and less readable code, potentially impacting the software development process. Similarly, DSU systems may not serve as a complete solution for system administrators if they still require approval or coordination before initiating updates, even without system downtime. Research into the usability of dynamic updating systems and avenues for improvement could potentially eliminate update timing concerns for administrators in the future.

## 11.5 Shifting Organizational Culture on Software Updates

In Section 10, we identified that organization management and policies can impact administrator actions, often impeding secure updating practices. A culture shift at organizations to recognize the importance of expedient updates (particularly for security issues) would help administrators perform their jobs more successfully. If end users and management do not readily accept that updates should be routinely applied, it becomes difficult to balance system maintenance and security with minimizing operational interruptions. Similarly, if organizations do not devote enough resources for administrators to adequately perform update tasks or have some oversight for security operations, security lapses can occur (e.g., Equifax [38]).

Resolutions to this problem are not straightforward. Existing recommendations such as NIST SP 800-40 [37] provide some guidance on organizational structures that promote updating. However, investigating how administrators deal with data breaches (similar to studies on end users facing breaches [52]) could provide insights into how to better facilitate practices that enable, not hinder, security, beyond solely relying on organizational security education. Such studies could also inform regulatory policies on security oversight. For instance, Equifax currently reports to 8 US states about their security overhaul [39]. The usable security community could offer insights into whether such audits fit into administrator workflows and improve security overall, or whether other policy approaches may better incentivize organizations to implement and prioritize security best practices.

# 12. CONCLUSION

System administrators play a vital role in securing machines on behalf of their organizations. One of their primary tasks is to manage the updates on numerous hosts to counter emergent vulnerabilities. However, prior work has paid less attention to how exactly they do so. In this paper, we examined how administrators manage software updates, determining five primary stages of updating and the various

considerations and actions associated with each stage. We identified pain points in administrator updating processes, such as when learning about updates, testing for and handling update-caused issues, deploying updates without causing operation disruptions, and dealing with organizational and management oversight. Based on our findings, we developed recommendations grounded in our results, and provided research directions for better support of administrators in keeping their hosts updated and secure.

## 13. ACKNOWLEDGMENTS

We thank our study participants, as well as Josefine Engel for running the pilot portion of our study. We also thank Serge Egelman, Katharina Krombolz, and Emanuel von Zezschwitz for meaningful discussions, and Noah Apthorpe for providing feedback on an earlier version of our paper. Finally, we thank our anonymous reviewers for providing constructive feedback. This work was supported in part by the National Science Foundation under awards CNS-1518921 and CNS-1619620.

## 14. REFERENCES

[1] Ansible. https://www.ansible.com/.

[2] Apache HTTPD Changelog. https://www.apache.org/dist/httpd/CHANGES_2.4.

[3] Chef. https://www.chef.io/chef/.

[4] Firefox Release Notes. https://www.mozilla.org/en-US/firefox/releases/.

[5] Ksplice. https://www.ksplice.com/.

[6] Puppet. https://puppet.com/.

[7] Reddit. https://www.reddit.com/.

[8] Spicework. https://www.spiceworks.com/.

[9] Terraform. https://www.terraform.io/.

[10] D. Armstrong, A. Gosling, J. Weinman, and T. Marteau. The Place of Inter-Rater Reliability in Qualitative Research: An Empirical Study. *Sociology*, 31(3):597–606, 1997.

[11] R. Barrett, E. Kandogan, P. P. Maglio, E. M. Haber, L. A. Takayama, and M. Prabaker. Field Studies of Computer System Administrators: Analysis of System Management Tools and Practices. In *ACM Conference on Computer Supported Cooperative Work (CSCW)*, 2004.

[12] S. Chiasson, P. van Oorschot, and R. Biddle. Even Experts Deserve Usable Security: Design Guidelines for Security Management Systems. In *SOUPS Workshop on Usable IT Security Management (USM)*, 2007.

[13] Cisco. Annual Security Report. https://www.cisco.com/web/offers/pdfs/cisco-asr-2015.pdf, 2015.

[14] O. Crameri, N. Knezevic, D. Kostic, R. Bianchini, and W. Zwaenepoel. Staged Deployment in Mirage, an Integrated Software Upgrade Testing and Distribution System. *SIGOPS Oper. Syst. Rev.*, 41(6):221–236, Oct. 2007.

[15] C. Dietrich, K. Krombholz, K. Borgolte, and T. Fiebig. Investigating System Operators' Perspective on Security Misconfigurations. In *ACM Conference on Computer and Communications Security (CCS)*, 2018.

[16] T. Duebendorfer and S. Frei. Why Silent Updates Boost Security. Technical report, ETH Zurich, 2009.

[17] Z. Durumeric, F. Li, J. Kasten, J. Amann, J. Beekman, M. Payer, N. Weaver, D. Adrian, V. Paxson, M. Bailey, and J. A. Halderman. The Matter of Heartbleed. In *ACM Internet Measurement Conference (IMC)*, 2014.

[18] S. Farhang, J. Weidman, M. M. Kamani, J. Grossklags, and P. Liu. Take It or Leave It: A Survey Study on Operating System Upgrade Practices. In *Annual Computer Security Applications Conference (ACSAC)*, 2018.

[19] A. Forget, S. Pearman, J. Thomas, A. Acquisti, N. Christin, L. F. Cranor, S. Egelman, M. Harbach, and R. Telang. Do or Do Not, There Is No Try: User Engagement May Not Improve Security Outcomes. In *USENIX Symposium on Usable Privacy and Security (SOUPS)*, 2016.

[20] C. Gkantsidis, T. Karagiannis, and M. Vojnovic. Planet Scale Software Updates. *ACM SIGCOMM CCR*, 36(4):423–434, Aug. 2006.

[21] M. Hicks and S. Nettles. Dynamic Software Updating. *ACM Transactions on Programming Languages and Systems*, 27(6):1049–1096, Nov. 2005.

[22] I. Ion, R. Reeder, and S. Consolvo. "...No one Can Hack My Mind": Comparing Expert and Non-Expert Security Practices. In *USENIX Symposium On Usable Privacy and Security (SOUPS)*, 2015.

[23] E. Kandogan, P. Maglio, E. Haber, and J. Bailey. *Taming Information Technology: Lessons from Studies of System Administrators*. Oxford University Press, 2012.

[24] S. Kraemer and P. Carayon. Human Errors and Violations in Computer and Information Security: The Viewpoint of Network Administrators and Security Specialists. *Applied Ergonomics*, 38(2):143 – 154, 2007.

[25] K. Krombholz, W. Mayer, M. Schmiedecker, and E. Weippl. "I Have No Idea What I'm Doing" - On the Usability of Deploying HTTPS. In *USENIX Security Symposium*, 2017.

[26] L. L. Kupper and K. B. Hafner. On Assessing Interrater Agreement for Multiple Attribute Responses. *Biometrics*, 45(3):957, Sept. 1989.

[27] F. Li, Z. Durumeric, J. Czyz, M. Karami, M. Bailey, D. McCoy, S. Savage, and V. Paxson. You've Got Vulnerability: Exploring Effective Vulnerability Notifications. In *USENIX Security Symposium*, 2016.

[28] F. Li, G. Ho, E. Kuan, Y. Niu, L. Ballard, K. Thomas, E. Bursztein, and V. Paxson. Remedying Web Hijacking: Notification Effectiveness and Webmaster Comprehension. In *World Wide Web Conference (WWW)*, 2016.

[29] H. Mann and D. Whitney. On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other. *Annals of Mathematical Statistics*, 18(1):50–60, 1947.

[30] A. Mathur and M. Chetty. Impact of User Characteristics on Attitudes Towards Automatic Mobile Application Updates. In *USENIX Symposium on Usable Privacy and Security (SOUPS)*, 2017.

[31] A. Mathur, J. Engel, S. Sobti, V. Chang, and M. Chetty. "They Keep Coming Back Like Zombies": Improving Software Updating Interfaces. In *USENIX Symposium on Usable Privacy and Security (SOUPS)*, 2016.

[32] A. Mathur, N. Malkin, M. Harbach, E. Peer, and S. Egelman. Quantifying Users' Beliefs about Software Updates. In *NDSS Workshop on Usable Security*, 2018.

[33] Microsoft. System Center Configuration Manager. https://www.microsoft.com/en-us/cloud-platform/system-center-configuration-manager.

[34] Microsoft. Windows Server Update Services. https://docs.microsoft.com/en-us/windows-server/administration/windows-server-update-services/get-started/windows-server-update-services-wsus.

[35] A. Nappa, R. Johnson, L. Bilge, J. Caballero, and T. Dumitraş. The Attack of the Clones: A Study of the Impact of Shared Code on Vulnerability Patching. In *IEEE Symposium on Security and Privacy (S&P)*, 2015.

[36] National Institute of Standards and Technology. National Vulnerability Database. https://nvd.nist.gov/.

[37] National Institute of Standards and Technology. Special Publication 800-40 Revision 3: Guide to Enterprise Patch Management Technologies. https://doi.org/10.6028/NIST.SP.800-40r3, 2013.

[38] L. H. Newman. Equifax Officially Has No Excuse. https://www.wired.com/story/equifax-breach-no-excuse/, September 2017.

[39] L. H. Newman. Equifax's Security Overhaul A Year After Its Epic Breach. https://www.wired.com/story/equifax-security-overhaul-year-after-breach/, July 2018.

[40] M. Oltrogge, Y. Acar, S. Dechand, M. Smith, and S. Fahl. To Pin or Not to Pin—Helping App Developers Bullet Proof Their TLS Connections. In *USENIX Security Symposium*, 2015.

[41] K. Rankin. Sysadmin 101: Patch Management. Linux Journal. https://www.linuxjournal.com/content/sysadmin-101-patch-management, 2017.

[42] I. Seidman. *Interviewing As Qualitative Research: A Guide for Researchers in Education and the Social Sciences*. Teachers college press, 2013.

[43] B. Stock, G. Pellegrino, F. Li, M. Backes, and C. Rossow. Didn't You Hear Me? Towards More Successful Web Vulnerability Notifications. In *Network and Distributed System Security Symposium (NDSS)*, 2018.

[44] B. Stock, G. Pellegrino, C. Rossow, M. Johns, and M. Backes. Hey, You Have a Problem: On the Feasibility of Large-Scale Web Vulnerability Notification. In *USENIX Security Symposium*, 2016.

[45] K. Vaniea, E. Rader, and R. Wash. Betrayed by Updates: How Negative Experiences Affect Future Security. In *ACM CHI Conference on Human Factors in Computing Systems (CHI)*, 2014.

[46] K. Vaniea and Y. Rashidi. Tales of Software Updates: The Process of Updating Software. In *ACM CHI Conference on Human Factors in Computing Systems (CHI)*, 2016.

[47] N. F. Velasquez and S. P. Weisband. Work Practices of System Administrators: Implications for Tool Design. In *ACM Symposium on Computer Human Interaction for Management of Information Technology (CHiMiT)*, 2008.

[48] N. F. Velasquez and S. P. Weisband. System Administrators As Broker Technicians. In *ACM Symposium on Computer Human Interaction for the Management of Information Technology (CHiMiT)*, 2009.

[49] R. Wash and E. Rader. Too Much Knowledge? Security Beliefs and Protective Behaviors Among United States Internet Users. In *USENIX Symposium On Usable Privacy and Security (SOUPS)*, 2015.

[50] R. Wash, E. Rader, K. Vaniea, and M. Rizor. Out of the Loop: How Automated Software Updates Cause Unintended Security Consequences. In *USENIX Symposium On Usable Privacy and Security (SOUPS)*, 2014.

[51] C. Weston, T. Gandell, J. Beauchamp, L. McAlpine, C. Wiseman, and C. Beauchamp. Analyzing Interview Data: The Development and Evolution of a Coding System. *Qualitative Sociology*, 24(3):381–400, 2001.

[52] Y. Zou, A. H. Mhaidli, A. McCall, and F. Schaub. "I've Got Nothing to Lose": Consumers' Risk Perceptions and Protective Actions after the Equifax Data Breach. In *USENIX Symposium on Usable Privacy and Security (SOUPS)*, 2018.

# APPENDIX

## A. PRELIMINARY PHASE - PILOT INTERVIEW QUESTIONS

Below we list the questions from our semi-structured pilot interviews (the preliminary phase of the study, as described in Section 3).

**Job responsibilities and processes**

1. Tell me more about your main job responsibilities (how does he/she keep machines up to date).
2. Tell me about any relationships you have with the vendors that develop the software updates for the programs your organization/employees depend on.
3. Can you walk me through your process of how you find out about an update?
4. Why do you find out about updates in this way?
5. How do you determine which updates to deploy on the machines you manage?
6. Tell me more about how this process differs for the types of machines you manage?
7. Why does your deployment process differ for different machines?
8. How does the process differ depending on who owns the machines, if at all?
9. Tell me more about how you install the software updates (manually, automatic, silent) you apply.
10. Why do you apply the software updates in this way?
11. Can you walk me through the process of testing whether an update will be compatible with the machines?
12. Why do you do this testing for the updates? Do you test all updates and why?

**Software Update Information**

1. Tell me about the information you currently receive when an update is available.
2. How do you usually receive this information?
3. Do you ever seek additional information about updates? Why or why not?
4. What are the main advantages of the current update information? Why?
5. What are the main disadvantages of the current update information? Why?
6. Which is the least important part of the current update information for you?
7. Which is the most important part of the current update information for you?

**Securing the Users**

1. Tell me about what you do to protect your users.
2. Tell me about what kinds of online hazards you are protecting them from.
3. Once an update is deployed how do you communicate the information to the end users?
4. What do you expect of the end users once the updates are released?

5. Can you tell me about the process of deciding what updates you can trust?

**Software Updates in General**

1. What updates are most important to you? Why?
2. What updates are least important? Why?
3. Tell me what cybersecurity means to you.
4. What are the most important things to consider to secure the network?
5. What are the least important things to consider to secure the network?
6. What are the main advantages of the current software updating process? Why?
7. What are the main disadvantages of the current software updating process? Why?
8. What changes would you want to make to software updates? Why?
9. Is there anything else you would like to tell us about how you manage software updates?

## B. PHASE ONE - SURVEY QUESTIONS

Below we list the questions from our survey (phase one of the study, as described in Section 3).

1. How old are you?
   (a) 18-25
   (b) 26-35
   (c) 36-45
   (d) 46-55
   (e) 56-65
   (f) Over 65
   (g) I do not wish to disclose
2. Which state do you live in?
3. What is your gender?
   (a) Male
   (b) Female
   (c) Other
4. What is your annual income?
   (a) Less than $25,000
   (b) $25,000 to $34,999
   (c) $35,000 to $49,999
   (d) $50,000 to $74,999
   (e) $75,000 to $99,999
   (f) $100,000 to $124,999
   (g) $125,000 to $149,999
   (h) $150,000 or more
5. What is your job title?
6. For how many years have you worked as a System Administrator in your current role?
7. For how many years have you worked as a System administrator before you entered your current role?
8. What is the highest level of education that you have completed?
   (a) 12th grade or less
   (b) High school degree or equivalent
   (c) Some college, no degree
   (d) Bachelor's degree
   (e) Master's degree
   (f) Other graduate degree
9. What was the subject area of your highest level of education (if above high school)?
10. What technical certifications, courses, or degrees have you completed, if any? You may paste entries from your resume or CV if you wish.
11. When did you complete these certifications or education? (Check all that apply)
    (a) Before I took up my current role
    (b) After I took up my current role
12. How have these technical certifications, courses, or degrees helped you complete your current role?
13. What is the industry of the organization that you work for?
14. How large is the organization that you work for?
    (a) $\leq$10 employees
    (b) 11 - 50 employees
    (c) 51 - 100 employees
    (d) 101 - 500 employees
    (e) 501-2000 employees
    (f) More than 2000 employees
15. What is the main purpose of the organization you work for?
16. How many machines/devices do you manage?
    (a) *Sliding scale between 0 and 1000+*
17. What type of machines/devices do you manage? (Check all that apply)
    (a) Laptops
    (b) Desktops
    (c) Servers
    (d) Mobile devices
    (e) Routers/network appliances such as firewall middleboxes
    (f) Embedded devices/ Internet of Things
    (g) Other: *free response*
18. What are the operating systems on the machines that you manage? (Check all that apply)
    (a) Mac
    (b) Windows
    (c) Linux
    (d) iOS
    (e) Android
    (f) Blackberry
    (g) ChromeOS
    (h) None
    (i) Other: *free response*
19. What is the predominant operating system, if any?
    (a) Mac
    (b) Windows
    (c) Linux
    (d) iOS
    (e) Android

(f) Blackberry
(g) ChromeOS
(h) Other: *free response*

20. What are these machines used for? (Check all that apply)
    (a) Education or training
    (b) Personal
    (c) Research
    (d) Servers
    (e) Work
    (f) Testing
    (g) Other: *free response*
21. Which of the following applies to the machines you manage? (Check all that apply)
    (a) The machines are used internally by the organization you work for
    (b) The machines are used externally by customers of the organization you work for
    (c) Other: *free response*
22. What updates are most important to you and why?
23. What updates are most important to your organization and why?
24. Are you solely responsible for updating the machines you manage?
    (a) Yes
    (b) No
    (c) Other: *free response*
25. How many updates do you run on the machines that you manage per week?
    (a) *Sliding scale between 1 and 500+*
26. How do you manage the updates across the machines/devices you manage? (Check all that apply)
    (a) I log into each system to perform updates
    (b) I use 3rd party software to manage the updates
    (c) I write programs to manage updates
    (d) I enable automatic updates
    (e) Other: *free response*
27. What type of updates do you install regularly? (Check all that apply)
    (a) Security updates
    (b) Non-security related updates
    (c) Other: *free response*
28. Select all of the security measures you take to protect your machines.
    (a) Firewall
    (b) Intrusion Detection System
    (c) Intrusion Prevention System
    (d) Antivirus System
    (e) Security updates
    (f) Different accounts with varying access (admin, regular, etc.)
    (g) Access codes/Passwords
    (h) Port scanners
    (i) Vulnerability testing
    (j) Backup and Disaster Recovery
    (k) Other: *free response*
29. How are the security measures you use deployed? (Check all that apply)
    (a) On the hosts
    (b) On the network
    (c) Other: *free response*
30. How do you find out about the updates you apply on the machines you manage? (Check all that apply)
    (a) Online forums
    (b) Security advisories
    (c) Blogs
    (d) News
    (e) Social media
    (f) RSS feeds
    (g) Professional mailing lists
    (h) Project mailing lists
    (i) Direct notification from vendor
    (j) Third-party service
    (k) When the software pops up a notification
    (l) Other: *free response*
31. When do you apply security updates? (Check all that apply)
    (a) As soon as they are released
    (b) After testing
    (c) On a regular cadence
    (d) After a specific amount of time since its release has elapsed
    (e) Applied automatically
    (f) Other: *free response*
32. What is the reason for applying updates in the frequency described above?
33. When do you apply non-security related updates? (Check all that apply)
    (a) As soon as they are released
    (b) After testing
    (c) On a regular cadence
    (d) After a specific amount of time since its release has elapsed
    (e) Applied automatically
    (f) Other: *free response*
34. What is the reason for applying non-security related updates in the frequency described above?
35. What kind of testing do you do with updates (if any), before applying them to the machines/devices you manage? Please explain why.
36. How frequently do you find an update to cause problems on the machines you manage?
    (a) Never
    (b) Rarely
    (c) Occasionally (every few update cycles)
    (d) Frequently (most update cycles)
37. How do you become aware of any problems caused by updates that you install?
38. What, if any, is your process for rolling back or undoing updates that cause problems on the machines you manage?
39. What aspects or steps in your update management process work well for you?
40. What aspects or steps in your update management process are most challenging to handle?
41. What would help you to better manage software updates for multiple machines?

## C. PHASE TWO - INTERVIEW QUESTIONS

Below we list the questions from our semi-structured interviews (phase two of the study, as described in Section 3).

**Job responsibilities and processes**

1. Tell me more about the company you work for?
2. Tell me more about your main job responsibilities (how does he/she keep machines up to date)
3. How long have you worked in your job?
4. Have you had any training in IT? If so, tell me more about that.
5. Have you had any training in security? If so, tell me more about that.

**Machines/Devices Managed**

1. Does your organization have any security related policies for their machines?
2. How many machines/devices do you manage?
3. What kinds of machines/devices do you manage?
4. What are these machines used for?
5. Who are these machines used by?

**Managing Software Updates for Multiple Machines**

1. Does your company have any policies on software updates for their machines?
2. How do you handle security for these machines?
3. How often do you update these machines? Does the frequency differ for different machines? If so, why?
4. Who do you have to notify about updates that you are applying? Why?
5. In an average week, how many hours do you spend dealing with software updates?
6. Can you walk me through your process of how you find out about an update?
7. What are the advantages of using this process?
8. What are the disadvantages of using this process?
9. How do you determine which updates to deploy on the machines you manage?
10. When do you apply updates for the machines you manage? Why?
11. What is your process for applying updates on the machines you manage?
12. Tell me more about how this process differs for the types of machines you manage.
13. Why does your deployment process differ for different machines?
14. How does the process differ depending on who owns the machines, if at all?
15. Tell me more about how you install the software updates (manual, automatic, silent) you apply.
16. Why do you apply the software updates in this way?
17. Do you use any tools/programs to help you manage updates on multiple devices? What are these tools? Why do you use them?
18. Do you test whether an update will be compatible with the machines you manage in any way? How so?
19. Why do you do this testing for the updates? Do you test all updates and why?
20. How do you track which updates different machines need?
21. Do you prioritize any particular type of updates for any machines? Why/why not?
22. How do you track how well updates have been installed on different machines?
23. If any update requires a restart, what is your process for managing the restart?
24. Do you have to notify anyone about updates that you have applied or are about to apply?

**Software Update Information**

1. Tell me about the information you currently receive when an update is available.
2. How do you usually receive this information?
3. Do you ever seek additional information about updates? Why or why not?
4. What are the main advantages of the current update information? Why?
5. What are the main disadvantages of the current update information? Why?
6. Which is the least important part of the current update information for you?
7. Which is the most important part of the current update information for you?
8. What improvements would you make to the information that is included with current updates?

**Securing the Users**

1. Who are the users that you manage machines for?
2. Tell me about what you do to protect your users.
3. Do you use any technical solutions to protect users?
4. Do you use any educational solutions for protecting your users?
5. Are these solutions driven by your own or company policy? Tell me more about that.
6. Tell me about what kinds of online hazards you are protecting them from.
7. Once an update is deployed how do you communicate the information to the end users?
8. What are your responsibilities for handling updates for your users?
9. What are the responsibilities of your users for handling updates?
10. Can you tell me about the process of deciding what updates you can trust?

**Software Updates in General**

1. What updates are most important to you? Why?
2. What updates are most important to your organization? Why?
3. How does your organizational policy influence how you manage updates if at all?

4. What updates are least important to you? Why?
5. What updates are least important to your organization? Why?
6. Tell me what cybersecurity means to you.
7. What are the most important things to consider to secure your machines?
8. What are the least important things to consider to secure your machines?
9. What are the main advantages of your current software updating process? Why?
10. What are the main disadvantages of your current software updating process? Why?
11. What would your ideal way to handle software updates be? Why? What changes would you want to make to software updates themselves? Why?
12. Is there anything else you would like to tell us about how you manage software updates?

# Communicating Device Confidence Level and Upcoming Re-Authentications in Continuous Authentication Systems on Mobile Devices

Lukas Mecke[1,2†], Sarah Delgado Rodriguez[2‡], Daniel Buschek[2†], Sarah Prange[1,3,2†], Florian Alt[3]
[1]*University of Applied Sciences Munich, Munich, Germany, {firstname.lastname}@hm.edu*
[2]*LMU Munich, Munich, Germany,† {firstname.lastname}@ifi.lmu.de, ‡S.Delgado@campus.lmu.de*
[3]*Bundeswehr University Munich, Munich, Germany, {firstname.lastname}@unibw.de*

## Abstract

Continuous implicit authentication mechanisms verify users over time. In case the device's confidence level (DCL) is too low, the user is prompted with a re-authentication request, which has been shown to annoy many users due to its unpredictable nature. We address this with a novel approach to enable users to anticipate the need for re-authentication with two indicators: (1) a *long term indicator* shows the current DCL and its development over time, and (2) a *short term indicator* announces that re-authentication is imminent. In both cases voluntary re-authentication allows the DCL to be raised and a device lock to be avoided. We tested the indicators in a four week field study (N=32). Our results show that both indicators were preferred over giving no indication and that importance and sensitivity of the interrupted task have a strong impact on user annoyance. Voluntary re-authentications were perceived as positive.

## 1 Introduction

Smart phones enable access to sensitive information, both on the device itself and in the cloud, that need to be protected. At the same time, traditional smart phone authentication is based on explicit authentication mechanisms, such as PINs, lock patterns, TouchID, and FaceUnlock. The use of such explicit mechanisms creates a considerable authentication overhead. Harbach et al. showed that smartphone users authenticate on average 47.8 times per day [16], spending 2.9% of their time on authentication.


*USENIX Symposium on Usable Privacy and Security (SOUPS)* 2019.
August 11–13, 2019, Santa Clara, CA, USA.

Researchers have proposed several methods to reduce authentication overhead, including time- or app-based approaches [7,18] as well as implicit authentication mechanisms that authenticate users based on their context [17,23] or their behaviour [6,11,12,24,27,28].

One caveat of such implicit authentication systems is that they can trigger explicit re-authentication; that is: asking users to confirm their identity via a second factor, in case the mechanism is unable to confirm the current user's identity [13,19,21]. Such re-authentication events are likely to interrupt other tasks and, hence, annoy users [20].

Reasons for this annoyance include the unpredictability of interruptions and the sensation of not being correctly informed about the current state of the implicit authentication system [2,9,20]. Moreover, users wish to influence the timing of the interruption in some way [2,22].

To address this, we propose (1) a long term indicator (*LT*), informing users about the current device confidence level (*DCL*) and thus enabling upcoming re-authentication to be anticipated, and (2) a short term indicator (*ST*), enabling users to finish their task. To avoid system-side locking of the device we (3) provide *voluntary* re-authentication (cf. Figure 1).

We investigated these indicators in a field study (N=32) where participants used them in everyday life. We found that people preferred our indicators to a system that interrupts them in an unpredictable way. Their perception strongly depended on the importance of the interrupted task. Voluntary re-authentication was perceived less annoying. Our research is complemented by deriving implications for future implicit authentication systems.

We contribute (1) novel designs to announce upcoming re-authentications and allow for voluntary re-authentication; (2) findings from a 4-week field study, testing the two indicators and their combinations; and (3) a set of implications for future implicit authentication mechanisms based on our findings.


Figure 1: We propose to use indicators to communicate both the current device confidence level (*DCL*) and the need for re-authentication for continuous implicit authentication systems on mobile devices: (1) a *long term* indicator illustrates the current *DCL* and its development over time via a task bar icon, and (2) a *short term* indicator announces an upcoming re-authentication via darkening the screen. Our system also allows for (3) *voluntary* re-authentication to avoid system-side locking of the device.

## 2 Underlying Use Cases

Implicit authentication has two major use cases: a) as an effortless, independent main authentication mechanism [19]; or b) as a second line of defence against unauthorised access to the private smartphone [21]. The first use case is particularly suitable for smartphone users that currently do not use any kind of authentication on their devices due to the required effort of explicit mechanisms. Hence, users would need to authenticate less frequently than with traditional explicit authentication approaches [16, 19]. The second use case provides an additional security barrier for devices which were already unlocked using an explicit mechanism [21].

In both cases, the reaction of the system to an unsuccessful authentication determines the provided security. An imminently triggered re-authentication prompt, as suggested by Khan et al. [19], promises to be one of the most secure approaches. But such interruptions could also be triggered by false rejects during an authorized usage and can therefore cause usability issues [20]. Some commercial products (e.g., Smart Lock[1]) instead keep the device unlocked and require re-authentication only after the session has ended. While this avoids interruption it also imposes a security risk, in case an attacker gets hold of the device within this time frame.

In this work we address systems that use interruptions to immediately lock the device as proposed by related work to minimise security risks. As previously shown, this can induce annoyance among users, which we aim to mitigate with appropriate indications to prepare users for upcoming re-authentications. Next, we discuss related work in this direction.

## 3 Related Work

### 3.1 Implicit Authentication

Many current authentication mechanisms rely on explicit authentication (i.e., recalling a secret or presenting a token or biometric feature [25]). The term *implicit authentication*[2], in contrast, describes the process by which a user is authenticated without requiring explicit interaction. In implicit authentication systems, the initial explicit authentication step to gain access to the device is replaced or complemented by a continuous evaluation of the users' identity that is reflected in a *device confidence level (DCL)*. Similar to a fallback in explicit authentication systems, an explicit so called *re-authentication* is required in case the device can not verify the user's identity.

Methods suggested for implicit authentication rely on the user's context [5, 17, 23, 26] and behavioural features. Examples include mechanisms that authenticate users based on gait recognition [12], continuous eye-tracking [24], or the users' tap or app-execution behaviour [6, 11, 27, 28].

There are several works pointing out the positive effects of implicit authentication. Hayashi et al. [17] found that implicit authentication could reduce explicit authentication by 68%. Riva et al. [26] report a decrease of 42%.

Several studies report on implicit authentication being perceived convenient and easier to use than traditional methods [8, 15, 20]. Finally, in a study by Crawford and Renaud [9] 90% of the participants indicated they would consider using implicit authentication and 73% felt it was more secure than authenticating explicitly.

### 3.2 Research on Re-Authentication

While implicit authentication is generally perceived positive and can indeed reduce authentication overhead, previous work found that the need for re-authentications can strongly disrupt those positive effects. Khan et al. [20] found that re-authentications, due to *false rejects (FR)* (i.e., cases in which the system rejected the legitimate user), were perceived annoying by 35% of their participants. This was due to both the unpredictable nature of the interruption and the need to switch the context for re-authentication. Another finding, also supported by the study of Crawford and Renaud [9], was that security barriers – like re-authentication – helped users to build a mental model of the system's security and thus led to a stronger perception of security.

[1] Smart Lock: https://support.google.com/android/answer/9075927?hl=en, last accessed June 24, 2019

[2] also called transparent or continuous authentication (e.g., [10]).

## 3.3 Interruptions

Work by Bailey et al. [4] found that interrupting users is perceived as rude and decreases task performance. They also found timing of an interruption to be highly important, as interrupted tasks were perceived as more difficult. Thus, they suggest using *attention manager* systems to detect phases of low memory load and schedule interruptions during these.

Adamczyk and Bailey [1] further investigated the impact of triggering interruptions at opportune moments. They were able to show that better timed interruptions are perceived as less annoying, less frustrating and more respectful. They also require less mental effort. Fischer et al. [14] aimed at identifying such opportune moments for interruptions with smartphones with the goal of identifying the best timing for delivering notifications. Although their participants did not clearly prefer the suggested interruptions after finishing a task compared to random interruptions, they found people attending faster to notifications in the task-dependent condition.

McFarlane [22] studied interruptions in general and found that making interruptions more predictable made them less annoying and had a positive effect on user performance in the interrupted task. He also found that letting users determine the moment of interruption made interruptions less annoying. Agarwal et al. [2] found similar results in their study. They tested different mechanisms to delay the re-authentication interrupt, using gradual dimming of the screen and transparent overlays to reduce context switch overhead and unpredictability of the interrupt. They found indications that participants were less annoyed when they could predict the interruption. Participants liked the introduced *grace period* (i.e., the delay of the re-authentication) and performance was increased as users tried to finish their tasks before the device was locked.

## 3.4 Implications of Related Work

From the insights in prior work we derive three opportunities for handling re-authentication interrupts in continuous authentication systems:

1. *Show current state*: Crawford and Renauds [9] found that users disliked the idea of a totally invisible authentication mechanism. Khan et al. [20] suggested indicating the current system status to address similar concerns voiced by participants of their study. This suggests that users' general desire for system feedback is particularly true for authentication as well.

2. *Announce interrupts*: Agarwal et al. [2] and McFarlanes et al. [22] found that predictable interruptions make users feel less annoyed.

3. *Delay interrupts*: Instantly locking the device when re-authentication is required can heavily disrupt the interaction flow [4]. Prior work showed that users liked having a *grace period* to finish their tasks in these situations [2].

# 4 Concept Development

In this section we report on the development process for our re-authentication concepts: We introduce design considerations revolving around *presentation strategy* and *integration with the smartphone*. These considerations provide the framing for a subsequent focus group in which participants brainstormed about specific designs. In the next section we describe our final concept for indicating upcoming re-authentications based on related work, our design considerations and our findings from the focus group.

## 4.1 Design Considerations

### 4.1.1 Presentation Strategy

From related work we derive two approaches for presenting a re-authentication indicator: *long-term* and *short-term*. We consider and investigate both.

**Long Term Indicator** To show the current state of the system, we consider a permanent indicator displaying the device confidence level (*DCL*) to show that the system is active. This also serves as a means to anticipate upcoming re-authentication.

**Short Term Indicator** To inform users about the imminent need for a re-authentication, we propose a short term indicator, granting a grace period.

### 4.1.2 Integration with the smartphone

The re-authentication indicator can be integrated with the smartphone in different ways: by means of static elements with the main purpose of permanently showing the current system status; by using dynamic elements, announcing an upcoming re-authentication request; or a combination of both approaches (hybrids).

**Static Elements** A well-suited static element on mobile devices is the task bar, as it is (with few exceptions) always shown. Possible elements are icons, percentages, progress bars or changes to the bar itself (e.g., changing colour) to indicate the current *DCL*.

**Dynamic Elements** On-screen dynamic elements include distortions of the screen content (e.g., darkening, desaturation, pixelation, etc. [2,3] ) or a notification. Off-screen elements include vibration, sound, the use of the flashlight, or the notification light.

**Hybrids** An element that can be used both statically and dynamically is a floating action button, overlaying screen content. Such buttons can show both *DCL* and upcoming re-authentication requests, either colour coded or in the form of e.g., a counter. In particular, a floating action button could also remain invisible and only (gradually) appear to announce a re-authentication.

#### 4.1.3 Freedom of Authentication

To address annoyance due to having to wait for the grace period to finish [2], we propose allowing explicit re-authentication at any time and in particular during the grace period.

### 4.2 Focus Group

The focus group served two purposes: (1) To collect novel design ideas for re-authentication concepts, focus group participants engaged in an open brainstorming session. (2) To understand users' preferences regarding the design opportunities, participants discussed several designs, covering different aspects of our considerations. We recruited five HCI students from our university (4 female, 1 male) for their expertise in interface design.

#### 4.2.1 Procedure

We first introduced participants to the concept of continuous implicit authentication and explained the terms 'device confidence level' (*DCL*) and 're-authentication'. Afterwards, we asked them to sketch ideas of how the current *DCL* and the need for re-authentication could be communicated to users. We provided print-outs of smartphone home-screens. Furthermore, we nudged them to think beyond visual cues. Following the sketching phase we asked them to present their ideas and discussed them. We then presented a set of our own indicator designs and asked participants to discuss those. Finally we asked participants to rank all designs (their own and our presented ones) and comment on why they chose a ranking.

#### 4.2.2 Focus Group Results

Results covered integration with the smart phone, visual design, modalities, and re-authentication mechanism.

Participants favoured approaches that subtly *integrate the indicator with the smartphone*. In particular, they felt that the indicator would optimally be placed in the task bar. Floating action buttons were perceived as too intrusive. Notifications received mixed opinions: While some participants argued that they were intrusive, others described them as the natural way the device would communicate announcements.

Regarding the *visual design*, participants suggested indicators gradually changing appearance (such as colour) to make users aware of diminishing *DCL*. Abrupt colour changes were considered too intrusive. A positively perceived idea was dimming the screen (similar to the method used in [2]).

Regarding *modality*, participants mentioned notifications and vibration to announce upcoming re-authentication.

As *re-authentication mechanism*, most participants mentioned biometric methods (fingerprint or face recognition) to make the process as smooth as possible. This is in line with feedback from participants in the study by Khan et al. [20].

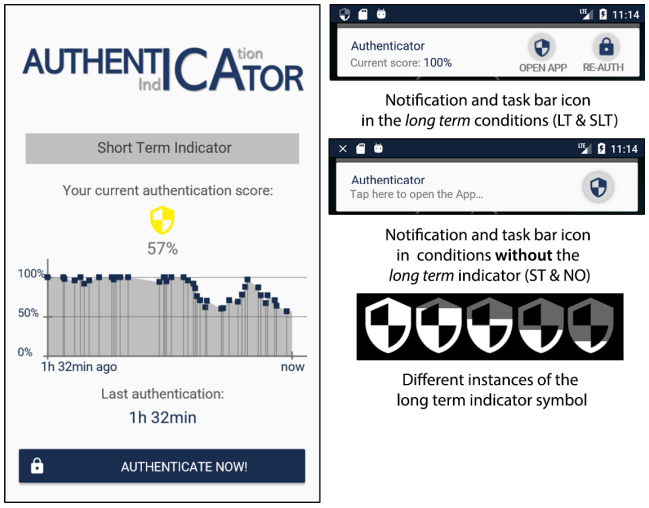



Figure 2: Different elements of the Authenticator app. Left: the main application with the device confidence level (*DCL*) visualised as a graph. Right: The notification and icon shown in the *long term* conditions (top), in the conditions without a *long term* indicator (middle) and the instances of the indicator symbol showing the current *DCL* in the task bar.

## 5 Authenticator

Based on the recommendations and suggestions both from related work and the focus group we built an android app, called *Authenticator*. The app simulates an implicit authentication system. It provides two different types of indicators that can be combined but also work independently.

### 5.1 Indicator Designs

Our prototype supports two indicators, namely a *short term* and a *long term* indicator.

#### 5.1.1 Long Term Indicator (LT)

To realise the long term indicator, our application places a permanent (non dismissable) notification in the task bar (cf. Figure 2 right top). As an icon we used a shield that gradually darkens in five steps, according to the *DCL* (cf. Figure 2 right bottom). In the notification, we displayed the current *DCL* value together with a button to open the control application and *re-authenticate voluntarily*. While we decided to permanently display the indicator in our study, it could also be implemented as an on-demand information source (comparable to e.g., battery level) to free up space in the task bar.

#### 5.1.2 Short Term Indicator (ST)

The short term indicator gradually darkens the screen once the *DCL* falls below 20% (Figure 1 centre). It is therefore only visible, when a re-authentication is imminent. To avoid

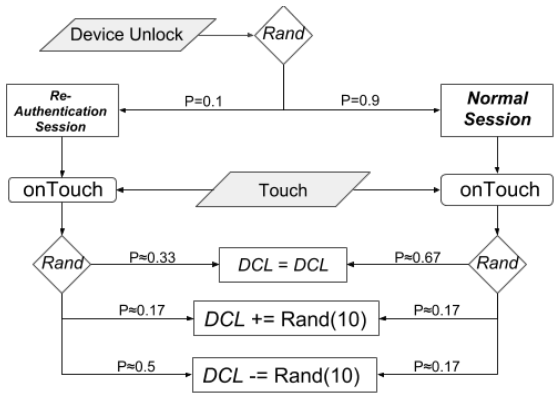



Figure 3: Schematic presentation of our simulated implicit authentication mechanism: Upon unlock of the device we determined (based on the desired false acceptance rate of 10%) whether a re-authentication should be triggered in this session (*re-authentication session*). The probabilities of user touches influencing the device confidence level (*DCL*) are altered accordingly; leading to decreases being more likely in *re-authentication sessions*. In *normal sessions* the *DCL* is more likely to remain stable.

annoyance through waiting for the grace period to end (cf. [2]), we display a notification as the dimming period begins. It shows a button to allow the user to *voluntarily re-authenticate* at any point within the grace period (Figure 2 right top).

In the study by Agarwal et al. [2] a duration of 4 seconds was chosen as shorter amounts did not allow for anticipation of the re-authentication and for longer duration testers had to wait too long for the re-authentication to appear. Due to the introduction of voluntary re-authentications the latter finding does not hold in our setting so we also explored longer grace periods. Through testing with five participants we determined a grace period duration of 8 seconds to be suitable. To address the remaining uncertainty we included a question about the desired length of the grace period in the final questionnaire.

## 5.2 Simulated Implicit Authentication

We followed related work and used a simulated system: Khan et al. [20] interrupted sessions after a random time period of between 5 and 30 seconds. Using a simulated system provides more control for our evaluation of the indicator concepts and helps to avoid differing false reject rates (e.g., due to hand posture) that might have an influence on the results [7, 9, 20]. We thus favoured a simulated system based on the number of touch interactions over a real implicit authentication system to keep conditions comparable. Following the medium-level false reject rate of 10% used in related work [20], our system triggers re-authentication in approximately one out of ten sessions[3]. To achieve this, we simulated DCL fluctuations as follows (cf. Figure 3):

[3] A session refers to the time between two unlocks.

### 5.2.1 Selection of Re-authentication Sessions

We flagged a session as a *re-authentication session* with a probability of 0.1 (to achieve 10% false rejects) upon unlocking the device. This flag influenced the random DCL fluctuations (see Figure 3) such that a re-authentication would likely appear in this session. For cases where sessions were too short for a re-authentication request to appear (i.e., the *DCL* did not fall below the threshold before the session ended), the flag would persist until a re-authentication was triggered. Depending on the flag being set or not, changes to the *DCL* were simulated differently, as explained next.

### 5.2.2 Alterations to the DCL

Depending on the chosen type of session (*re-authentication* or *normal*) the goal was to either decrease *DCL* or keep it stable while adding some fluctuation to make the results more believable. Each touch by the user had a chance to either trigger a change to the *DCL* (0.67 if it was a *re-authentication session*, 0.33 in a *normal session*) or leave it unchanged (with inverse probability accordingly). For *re-authentication sessions*, a decrease of the *DCL* was more likely (0.5) in comparison to increases (0.17). In *normal sessions* the probability for decreases and increases was equal at 0.17 (compare Figure 3 for an overview of the whole process). Both decreases and increases to the *DCL* could trigger a random change between 1% and 10%. Decreases resulting in a *DCL* below 20% were only executed in *re-authentication sessions*.

All probabilities were determined through a pre-study with five testers so as to create fluctuation of the *DCL* that seemed natural. A re-authentication was triggered as the *DCL* fell below 20% and completing a re-authentication reset the *DCL* to 100%. Re-authentication was suspended during calls.

### 5.2.3 Usage

Using this method we achieved an actual false reject rate of 7.65% in our 4-week field study. The deviation from the goal (10%) is a result of sessions that were too short to trigger a re-authentication. While we forced the next session to be a re-authentication session in those cases as described above, we did not adjust probabilities afterwards to mitigate effects on the overall false reject rate.

## 5.3 Re-Authentication

Voluntary re-authentication was possible using the control application (Figure 2 left) or one of the notifications tied to the indicators (Figure 2 right), i.e., the permanent notification or the notification displayed during the grace period. Information about the current *DCL* was provided by the permanent notification icon (discretised), the permanent notification, and the control application. The latter additionally featured a graph, displaying the history of the *DCL* over time (Figure 2 left).

The *re-authentication process* itself was implemented by locking the device and, hence, forcing the user to authenticate by using their default unlock mechanism. Due to technical restrictions it was not possible to offer biometric methods for re-authentication as Android requires using the backup authentication scheme in cases where the device is locked by an app. Using those methods was still possible for normal locks, i.e., locks that were not triggered by our app.

## 6 Evaluation

Our evaluation was guided by the these research questions:

**Q1 –** *Can indicators reduce annoyance caused by unpredictable re-authentication requests?* We hypothesise this to hold true due to results from related work [2, 22].

**Q2 –** *Are there other factors influencing annoyance caused by re-authentication requests?* We propose location, task and importance and sensitivity of the interrupted task as possible factors.

**Q3 –** *Do indicators nudge users to voluntarily re-authenticate?* We expected an increasing number of voluntary re-authentications for short term (due to the option to re-authenticate during the grace period) and long term indication (due to the added feedback from the task bar symbol and the graph visualisation of the *DCL*).

**Q4 –** *How do users perceive and respond to the introduction of voluntary re-authentication?* We expected users to like this feature, as prior work showed that letting users determine the interruption time reduced annoyance [22].

### 6.1 Study Design

To answer our research questions we conducted a field study (N=32). The study employed a within-subject design. Participants tested a set of four conditions for one week each, resulting in a total study length of four weeks. The order of conditions was counterbalanced.

1. (**NO**) ***No** Indication*: Our (simulated) implicit authentication scheme runs transparently in the background. Re-authentication is requested without prior indication, which resembles the current practical standard. Voluntary re-authentication is only possible from the control app, but not from notifications.

2. (**ST**) ***S**hort **T**erm*: Only the *short term* indicator is shown. Voluntary re-authentication is possible from the control app and the notification triggered with the grace period.

3. (**LT**) ***L**ong **T**erm*: Only the *long term* indicator is shown. Voluntary re-authentication is possible from the control app and the permanent notification.

4. (**SLT**) ***S**hort & **L**ong **T**erm*: Both indicators are present. All options for voluntary re-authentication are possible.

Note how both *NO* and *ST* can serve as baselines here. The *NO* condition, i.e., locking the device without giving indication, is the current *practical* state of the art and thus a natural baseline. Furthermore our *ST* condition is based on the best performing method from the study by Agarwal et al. [2] (including their recommended change of allowing for re-authentication during the grace period). As such, *ST* serves as a baseline for the best currently known scheme for indicating re-authentications.

### 6.2 Procedure

We recruited participants through a University mailing list and via social media. They were asked to sign a consent form and install our app from the Google Play Store, using an installation guide we provided on a dedicated website. This website also provided additional information about all study conditions and answers to frequently asked questions.

Participants had to *use the application* for four weeks with conditions automatically switching each week. They used their phones as usual with occasional interruptions by our system and a maximum of three (dismissible) *experience sampling questionnaires* per day after successful re-authentication. After each condition switch, we asked participants to fill a *weekly questionnaire* about their experience. After all conditions we concluded with a *final questionnaire*.

After four weeks, participants could uninstall the app and we invited them to participate for a *final semi-structured interview* to collect qualitative feedback (in person or via telephone). Participants received €20, plus €5 if they participated in the interview.

### 6.3 Collected Data

We collected *usage data* on participants' devices, including executed apps, and aggregated touch interactions, unlocks, and re-authentications. Collected data was stored on the device and transferred to our server once per day.

The *experience sampling questionnaires* asked for current location and interrupted task. We also asked if the interrupted task was perceived as sensitive and important and if the interruption was perceived as annoying.

In our *weekly questionnaires*, participants rated on a 5-point Likert scale if they felt rewarded by an increasing *DCL*, if they felt motivated to re-authenticate voluntarily, and if they perceived the system as obstructive, annoying, and easy to use. We also asked for free feedback on what they liked and disliked about the current indicator and the system in general.

In the *final questionnaire* we asked participants to rank the four conditions and explain their decision. In particular, we asked which features of the first and last choices contributed

| | | |
|---|---|---|
| **Gender** | 14 (44%) | Female |
| | 18 (56%) | Male |
| **Mean Age** | 28.3 | |
| **Occupation** | 2 (6%) | Homemaker or retiree |
| | 8 (25%) | Working |
| | 22 (69%) | Student |
| **Primary Unlock Mechanism** | 1 (3%) | Password |
| | 2 (6%) | PIN |
| | 2 (6%) | Face Recognition |
| | 6 (19%) | Pattern |
| | 21 (66%) | Fingerprint |
| **Secondary Unlock Mechanism** | 3 (9%) | Password |
| | 8 (25%) | PIN |
| | 10 (31%) | Pattern |
| | 11 (34%) | None |
| **smart phone usage (mean)** | 52.7 | Estimated daily unlocks |
| | 3.6 | Estimated daily usage (h) |

Table 1: Demographics of the participants of our four week field study (N=32).

to their decision. For the specific indicators, we asked participants whether they would modify the duration of the grace period, if they were stressed due to the grace period, and if the long term indicator helped predicting re-authentications.

Furthermore, participants rated several statements on a 5-point Likert scale: Did they like the system, were they annoyed by the vibration or notification (*ST*), did they feel that the system influenced their behaviour, and did any bugs influence the system performance? Similarly, we asked participants if the experience sampling was annoying, and if it influenced their behaviour or the perception of the system.

Moreover, we asked if participants had read the introduction on the website and watched the introductory video we provided, if they had previous knowledge about implicit authentication, and if they had looked up app functionality or how implicit authentication worked in general on our website or other sources. Finally, we asked if they always locked their phone after use, if they thought re-authentication interrupts were more annoying than traditional authentication, and if they would consider using implicit authentication.

In the *final interview*, we asked participants to share their experiences with the systems guided by a few questions.

## 6.4 Participants

We recruited 36 participants. Four were excluded since their data was not properly transferred to our server. The remaining 32 people had a mean age of 28 years (18 male and 14 female; Table 1). Three participants did not submit a final questionnaire, resulting in a reduced set of 29 answers for these questions. For practical reasons we conducted the study in two runs (i.e., not all participated in parallel).

All but two participants partially agreed (n=7) or agreed (n=23) that the restriction of access to their smartphone (authentication) was important (5-point Likert scale). Participants self reported their technical knowledge as high (median=4).

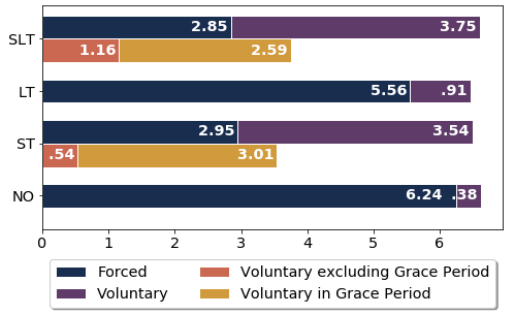



Figure 4: Average daily re-authentications by condition. Re-authentications are divided in voluntary and forced re-authentications and voluntary re-authentications are again subdivided in re-authentications during and excluding the grace period (where applicable).

## 6.5 Study Limitations

As participation were self-selected, our sample may not represent the general population. Our simulation might differ from the dynamics when using real implicit authentication systems. Moreover, our prototype added re-authentication on top, whereas a real system could in turn remove the initial device unlock authentication. This might have negatively affected participants' perception of our system. However, the goal was not to evaluate the general concept of implicit authentication itself but indicators for re-authentication.

# 7 Results

In the following report, quantitative results were tested for significance using repeated measures ANOVA with Greenhouse-Geisser correction and Bonferoni post-hoc tests. Ordinal results were tested using a Friedman test with Conover's post-hoc tests. We report significance at the level of $p < 0.05$. No effects of ordering were observed.

## 7.1 Usage Data

Over the course of the four week field study we observed a total of about 3.6 million touches and about 74.200 unlocks (average 84.7 unlocks per day and user) of which 5679 (7.65%) were re-authentications (1910 were voluntary, of which 646 were outside of the grace period).

The *average number of daily re-authentications* per condition is shown in Figure 4. We found no effect of the indicators on the average number of daily re-authentications. However, we found a significant difference for the average number of daily *voluntary* re-authentications ($F(1.95, 60.44)=14.75$, $p<.001$, $\eta^2 =0.322$). Post-hoc tests revealed significantly more voluntary re-authentications for all indicators ($p<.04$) compared to none (*NO*); and also significantly more for *ST* ($p=.001$) and *SLT* ($p=.003$) compared to *LT*.

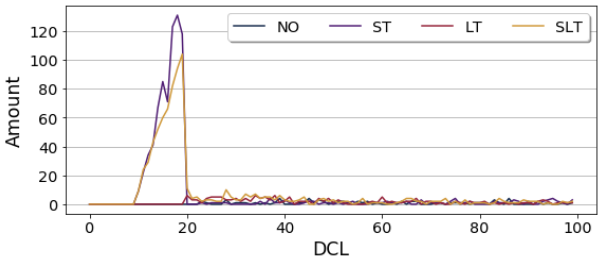



Figure 5: Distribution of *DCL* at voluntary re-authentication. There are no re-authentications below 20% for *NO* and *LT* as they had no grace period but instantly locked the device.

We also analysed re-authentications *excluding* those in the grace period, since these are arguably not strictly voluntary: We found a significant difference for relative daily voluntary use, that is, the ratio of voluntary to all re-authentications ($F(2.82, 84.53)=59.09$, $p<.001$, $\eta^2=0.165$). Post-hoc tests revealed significantly higher relative voluntary re-authentication for both *LT* ($p=.014$, Mean=14.56%) and *SLT* ($p=.008$, Mean=17.63%), compared to *NO* (Mean=5.67%). Relative voluntary re-authentications *during* the grace period were significantly higher ($F(1.0, 30.0)=5.01$, $p=.032$, $\eta^2=0.144$) for *ST* (Mean=47.49%) than for *SLT* (Mean=38.93%).

In 49.6% of cases, participants re-authenticated *before* the grace period was over, that is, they did not wait for system-triggered re-authentication (Mn=3.29s, SD=1.46). Outside of the grace period, there was no particular *DCL* at which people preferred to voluntarily re-authenticate (Figure 5).

In summary, we did not observe an effect of the indicators on the *total* average daily re-authentications. However, *voluntary* re-authentications were more common when using indicators. This can be mainly attributed to re-authentications *outside* the grace period for conditions including the long term indicator and re-authentications *during* the grace period for conditions using the short term indicator.

## 7.2 Experience Sampling

### 7.2.1 General Results

We collected 1557 answers for the experience sampling questionnaires. On a 5-point Likert scale, annoyance was rated neutral *over all conditions* (Median=3). The statements that the interrupted task was sensitive and that the interrupted task was important were also rated neutral (both Median=3). We could not find a significant impact of indicators on any rating.

Regarding the *authentication context*, participants most frequently reported "at home" for the *place* where they were interrupted, followed by transit and work. The most frequent *tasks* that were interrupted were chatting, reading, searching for information, "nothing"[4] and writing. This aligns with our logged data about the interrupted apps.

[4]This includes both cases where participants actually did nothing in particular or were not interrupted, as the re-authentication was voluntary.

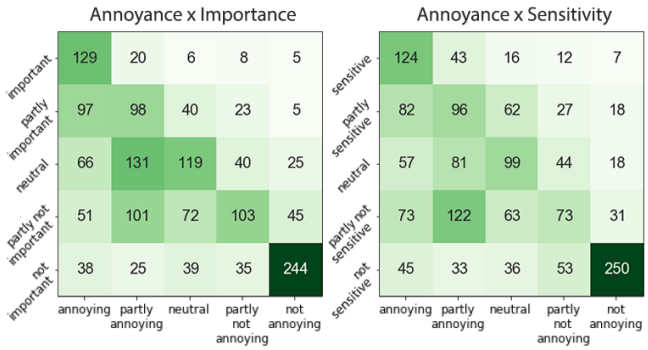



Figure 6: Frequencies of reported annoyance by importance of the interrupted task (left) and by sensitivity of the interrupted task (right). Colour encodes the shown counts.

### 7.2.2 Annoyance

We found significant positive (Spearman) correlations between perceived annoyance and importance of the interrupted task ($r_s=0.569$, $p<.001$) and between perceived annoyance and sensitivity of the interrupted task ($r_s=0.489$, $p<.001$), see Figure 6. We could not find effects of the day of the week or the day since the specific condition started.

The annoyance of voluntary re-authentication was perceived neutral (n=273, Median=3), similar to forced re-authentication (n=1277, Median=3). The degree to which people were annoyed by voluntary re-authentication did not significantly differ based on whether it happened during (n=76, Median=3.5) or outside of the grace period (n=136, Median=3). Voluntary re-authentication was labelled as such in the experience sampling in only 18.3% of the cases.

When comparing annoyance for the most frequently reported tasks in the experience sampling, a Friedman test revealed a significant effect of task on annoyance through re-authentication ($\chi^2(5)=36.16$, $p<.001$, $W=0.604$). Conover's post-hoc tests found that the interruption of the task "voluntary/nothing" was perceived as less annoying (Median=1) when compared to chatting ($p<.001$, Median=4), reading ($p=.002$, Median=3), searching for information ($p<.001$, Median=4), writing ($p<.001$, Median=4) and all other tasks ($p<.001$, Median=4).

In summary, we found that the annoyance caused by an interruption was influenced by a) the sensitivity of the data accessed during the interrupted task, b) the importance of the interrupted task, and c) by the task itself, as the reported task "voluntary/nothing" was perceived as less annoying.

## 7.3 Weekly Questionnaires

### 7.3.1 Voluntary re-authentications

For the weekly questionnaires we found significant differences for the motivation to voluntarily re-authenticate

($\chi^2(3)=10.05$, p=.018, W=0.498) and the feeling of reward by an increased *DCL* after re-authentication ($\chi^2(3)=21.74$, p<.001, W=0.618) with regards to the different indicators. Post-hoc analysis revealed that for *SLT* (Median=3) participants felt significantly more motivated to voluntarily re-authenticate than for *NO* (Median=1, p=.009). For all conditions using an indicator participants felt significantly more rewarded (Median-ST=2, Median-LT=2, Median-SLT=3) than in the *NO* condition (Median=1, p<.02). We found no significant differences on *perceived annoyance* of the system.

Thus, while we cannot provide evidence for a general effect of our indicators on the annoyance, we did find a positive influence of the long term indicator on the motivation to voluntarily re-authenticate. The feeling of being rewarded for re-authentication by the increased *DCL* was also significantly higher for the conditions including the long term indicator.

#### 7.3.2 Perception of Indicators

Participants liked about the indicators that interruptions were less sudden compared to no indication (mentioned by 22 people) and that the *DCL* was visible at any time for the conditions with a long term indicator. In the *NO* condition, participants liked that re-authentication was fast (9 mentions). The gradual darkening was positively mentioned by ten participants for *ST* and by eight for *SLT*.

Interrupts were perceived as sudden by fifteen participants in the *NO* condition and by ten, four and three participants in the *LT*, *ST* and *SLT* conditions, respectively. Seven participants reported they overlooked the *DCL* visualization in the *LT* condition. Interrupts were in general perceived as annoying in all conditions (mentioned by 10, 9, 7 and 8 participants for the *NO, ST, LT* and *SLT* conditions, respectively).

### 7.4 Final Questionnaire

#### 7.4.1 Ranking

In the final questionnaire, participants were asked to rate their experience with the system in general. The *overall ranking* of the different conditions (Figure 7) reveals that the combination of both *long term* and *short term* was preferred. No indication (*NO*) was ranked last. Long term (*LT*) and short term (*ST*) ranked second and third. Based on the open questions, the following reasons contributed to their choice: Sixteen participants stated to not like the sudden interruptions without indication. The combination of both short and long term (*SLT*) was particularly liked for the best overall overview and control and the continuous visualization of the *DCL* (10 and 9 mentions).

#### 7.4.2 General Perception

As a response to our Likert scale questions, participants did not find vibration and notifications particularly annoying

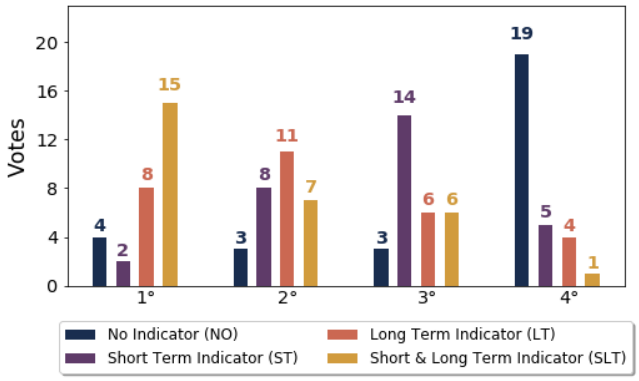



Figure 7: Participants' ranking of the different indicators. The combination of long- and short term indicator was the most preferred method while no indication was least preferred.

(Median=2). They felt neutral towards being stressed by the dimming during the grace period (Median=3). The long term taskbar symbol was considered to be helpful (Median=4) to predict re-authentications.

Participants remained neutral (Median=3) towards a possible influence of the system on their behaviour. They partly liked the design (Median=4) and partly disagreed to being negatively influenced by bugs (Median=2). They felt neutral (Median=3) about the experience sampling being annoying or influencing their behaviour or perception.

No one had profound knowledge about implicit authentication before the study nor did they review implicit authentication from other sources than the material provided by us (Median=1). There was general agreement on having read the introduction on the website and having watched the whole introductory video (Median=5).

In general, participants agreed to always locking their device (Median=5) and to authentication interrupts being more annoying than traditional authentication up front (Median=5). Regarding whether they would use the concept of implicit authentication in general, participants remained neutral (Median=3; 10 agreed or partly agreed, 5 neutral, and 14 disagreed or partly disagreed).

Finally, people would have liked a slightly longer grace period. On average they suggested 10.14 s (range 2 s–60 s).

## 8 Discussion & Implications

### 8.1 Importance & Sensitivity

While we did not find a significant effect of indicators on perceived annoyance via experience sampling, we gained related evidence and insights: We found a significant impact of *sensitivity* and *importance* of an interrupted task on the perceived annoyance. This was also pointed out in the final in-

terviews where five of the eight participants found the system interrupting an important or stressful task to be a particularly negative event:

> *I remember when I had to make a really important call and my screen was locked before I could do it. I had to answer the feedback, too, before I could finally call. Then, it was really annoying, but usually the interrupts were no problem.*

As a key insight, the situations in which participants perceived interrupts as annoying were also those that they rated as sensitive, hence, those that would require increased protection when relying on a real implicit authentication system. It might be possible that users were biased as they knew their phone was protected by their primary locking mechanism anyway in this study. Nevertheless, we believe that this topic should be investigated further.

## 8.2 Voluntary Re-Authentication

In contrast to related work on general interruptions [22], we could not find a positive effect of deciding when to re-authenticate on reducing annoyance. For the grace period, one explanation is that participants might not have perceived the option to re-authenticate as voluntary (as re-authentication was inevitable). More generally, our results on importance, sensitivity, and interrupted tasks all point towards the conclusion that for our participants annoyance was mostly determined by the interrupted activity and not by whether it was voluntary or not.

Nevertheless, voluntary re-authentications were mentioned as positive in open comments and the interviews, and indeed accounted for a considerable proportion of 33.6% of re-authentications (11.4% excluding grace period). Moreover, users felt significantly more motivated to re-authenticate for the combined short and long term indicator. All indicators also resulted in significantly more common use of voluntary re-authentications.

Hence, a promising approach to reduce user annoyance might be to investigate concepts that provide options for users to voluntarily re-authenticate with awareness of current activities. For instance, one person suggested to allow for voluntary re-authentication when opening an app, which often coincides with the beginning of a new activity.

## 8.3 Grace Period

We received mixed feedback on the grace period. Many participants liked it, in particular the more predictable nature of the interruption. For example, one participant said:

> *The more sudden the interruption happened, the more annoyed I felt about it. Surprisingly, it did not depend so much on the frequency of the interrupts. It only depended on the announcement.*

However, some participants complained that they could not use the grace period to its full extent due to light conditions and wished for a longer duration. Others used our introduced option to voluntarily re-authenticate before the device was locked. In general the desired length was very different amongst the participants which implies that an option to customise this (as also suggested by Agarwal et al. [2]) might indeed be promising for future work. We also believe that there is an impact of the personal *usability-security trade-off*, as having a (longer) grace period also implies a security risk in cases where an attacker would get hold of the device. Steps to address this might be, e.g., adapting the length of the grace period to the derivative of the *DCL* (i.e., strength of change in system confidence) or the importance of the interrupted app.

In general we see the approach of gradually dimming the screen only as a first step. Moreover, as proposed by participants of our focus group, future systems could, for example, use biometrics for re-authentication. In this case, dimming the screen could be an indicator for the user to present their face to the camera or quickly put the finger on a fingerprint scanner and thus avoid a full context switch.

## 8.4 Interruptions

Based on the previously discussed results, we present three recommended aspects to consider with regard to scheduling re-authentication interrupts.

1. **Sensitivity of the task:** If the user is accessing non-sensitive data (e.g., while reading a book), an upcoming re-authentication could be delayed or triggered when the task is finished, as suggested by related work [1, 4] and done in practice[5]. However, while accessing sensitive data (e.g., banking app), re-authentication should be triggered instantly to restrict further access.

2. **Importance of the task:** As users found interruptions of important tasks particularly annoying, selectively delaying such interruptions could improve users' experience with the system. This assumption is further supported by Adamczyk and Bailey [1, 4].

3. **Recent changes in confidence:** Changes in device confidence level (*DCL*) over time may be used as an indication for the necessity of an immediate interruption. While a sudden decrease in confidence most likely corresponds to an intruder taking hold of the device, a slow decrease is more likely to be caused by natural variations in the legitimate user's behaviour. However, those assumptions are, as of now, speculative and further research with a functioning implicit authentication system is necessary to verify this hypothesis.

[5] e.g., Smart Lock: https://support.google.com/android/answer/9075927?hl=en, last accessed June 24, 2019

The focus of our work was on interruptions caused by a continuous authentication system. Some lessons learned may generalise to other interruptions, such as notifications. A further factor to consider in that case is the importance of the interruption itself – which we assumed to be high for implicit authentication due to the security risk.

### 8.5 System Design

For our study we introduced a novel method to more realistically simulate an implicit authentication system. Our approach extended previous approaches (e.g., Khan et al. [20]) and made some of our evaluations, like the *long term* indicator, possible in the first place. We believe this to be a valuable step to enable future evaluations but also acknowledge that using our system had limitations. In particular, as the system was touch-based we introduced a bias towards interrupting tasks that used many touches, such as writing, whereas very short interactions were interrupted less. One way to address this would be to track the current app and schedule interrupts to distribute re-authentication request equally over the different tasks. Due to our use of a simulated system we were also not able to remove the primary unlocking mechanism, as this would have left participants unprotected.

However, our results from the final questionnaire suggest that neither the system itself nor the introduced experience sampling had a major effect on participants' perception or behaviour. Furthermore, vibration feedback and notifications were not perceived as annoying, and the overall design was rated as very positive.

### 8.6 Adoption of implicit authentication

Our participants remained neutral towards using implicit authentication and only 10 of 29 agreed or partially agreed to wanting to use it. This contrasts results of previous studies: Crawford and Renaud [9] report 90% of their participants to be interested in adopting implicit authentication. Participants also generally agreed that re-authentication was more annoying than unlocking up front.

Possible reasons could be that users underestimate the actual number of authentications they perform (on average by 38% in our study) and the accompanying benefit of implicit authentication. Other explanations include authentication overhead of a simulated system, or habituation to users' traditional unlocking methods. On the other hand, studies from related work were a lot shorter (several lab studies [2, 9, 26] and shorter field studies [20]) and thus user perception in our study developed over a longer period of time (e.g., we potentially observed a lower novelty effect). Moreover, effortless fingerprint authentication in particular has become an established method in the years between some of the earlier related work and our study, potentially shifting users' views.

As a next step we suggest evaluations with a functional implicit authentication system for a more realistic scenario. In cases where such a system cannot robustly provide sufficient security, conducting the study with users that do not lock their phones anyway might be an option. Targeting this user group has also been suggested as a mayor application area for implicit authentication in related work [19, 29].

### 8.7 Research Questions

Regarding our initial research questions we found all our indicators being preferred to no authentication.

We found no effect of indicators on annoyance. Annoyance was rather determined by the interrupted activity (Q1). We found sensibility, importance, and the specific interrupted task to be further factors influencing the perceived annoyance of interrupts (Q2). We also found all indicators to have a positive effect on the use of voluntary re-authentications (Q3). Finally, we found that users felt particularly motivated to voluntarily re-authenticate by combined short and long term indication. They overall perceived voluntary re-authentication as positive and used it to a considerable extent (Q4).

## 9 Conclusion

Motivated by previous work finding unpredictability of re-authentication requests in implicit authentication systems a source of annoyance we introduced and evaluated two indicator designs. Those included a *long term* indicator constantly showing the system confidence and a *short term* indicator announcing imminent re-authentications and giving users a grace period to finish their tasks. We also introduced *voluntary re-authentications* to allow users to re-authenticate at any time and skip the grace period if desired.

From the results of our four week field study (N=32), we found that both indicators were preferred to having no indication. We also found our newly introduced conditions to be preferred over the indicator motivated by previous work and that importance and sensitivity of the interrupted task are further influencing factors on user annoyance.

We hope for our insights to provide fertile ground for designers of future implicit authentication systems with the goal of making them as usable as possible and further support the endeavour of blending authentication seamlessly with the way that users interact.

## 10 Acknowledgements

Work on this project was partially funded by the Bavarian State Ministry of Education, Science and the Arts in the framework of the Centre Digitisation.Bavaria (ZD.B). This research was supported by the Deutsche Forschungsgemeinschaft (DFG), Grant No.: AL 1899/2-1.

## References


[1] Piotr D Adamczyk and Brian P Bailey. If not now, when?: the effects of interruption at different moments within task execution. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 271–278. ACM, 2004.

[2] Lalit Agarwal, Hassan Khan, and Urs Hengartner. Ask me again but don't annoy me: Evaluating re-authentication strategies for smartphones. In *Symposium on Usable Privacy and Security (SOUPS)*, 2016.

[3] Florian Alt, Andreas Bulling, Gino Gravanis, and Daniel Buschek. Gravityspot: guiding users in front of public displays using on-screen visual cues. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pages 47–56. ACM, 2015.

[4] Brian P Bailey, Joseph A Konstan, and John V Carlis. The effects of interruptions on task performance, annoyance, and anxiety in the user interface. In *Interact*, volume 1, pages 593–601, 2001.

[5] Jakob E Bardram, Rasmus E Kjær, and Michael Ø Pedersen. Context-aware user authentication–supporting proximity-based login in pervasive computing. In *International Conference on Ubiquitous Computing*, pages 107–123. Springer, 2003.

[6] Attaullah Buriro, Bruno Crispo, Filippo Del Frari, and Konrad Wrona. Touchstroke: smartphone user authentication based on touch-typing biometrics. In *International Conference on Image Analysis and Processing*, pages 27–34. Springer, 2015.

[7] Daniel Buschek, Fabian Hartmann, Emanuel Von Zezschwitz, Alexander De Luca, and Florian Alt. Snapapp: Reducing authentication overhead with a time-constrained fast unlock option. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 3736–3747. ACM, 2016.

[8] Nathan Clarke, Sevasti Karatzouni, and Steven Furnell. Flexible and transparent user authentication for mobile devices. In *IFIP International Information Security Conference*, pages 1–12. Springer, 2009.

[9] Heather Crawford and Karen Renaud. Understanding user perceptions of transparent authentication on a mobile device. *Journal of Trust Management*, 1(1):7, 2014.

[10] Heather Crawford, Karen Renaud, and Tim Storer. A framework for continuous, transparent mobile device authentication. *Computers & Security*, 39:127–136, 2013.

[11] Alexander De Luca, Alina Hang, Frederik Brudy, Christian Lindner, and Heinrich Hussmann. Touch me once and i know it's you!: implicit authentication based on touch screen patterns. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 987–996. ACM, 2012.

[12] Mohammad Omar Derawi, Claudia Nickel, Patrick Bours, and Christoph Busch. Unobtrusive user-authentication on mobile phones using biometric gait recognition. In *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2010 Sixth International Conference on*, pages 306–311. IEEE, 2010.

[13] Tao Feng, Ziyi Liu, Kyeong-An Kwon, Weidong Shi, Bogdan Carbunar, Yifei Jiang, and Nhung Nguyen. Continuous mobile authentication using touchscreen gestures. In *Homeland Security (HST), 2012 IEEE Conference on Technologies for*, pages 451–456. Citeseer, 2012.

[14] Joel E Fischer, Chris Greenhalgh, and Steve Benford. Investigating episodes of mobile phone activity as indicators of opportune moments to deliver notifications. In *Proceedings of the 13th international conference on human computer interaction with mobile devices and services*, pages 181–190. ACM, 2011.

[15] Cristiano Giuffrida, Kamil Majdanik, Mauro Conti, and Herbert Bos. I sensed it was you: authenticating mobile users with sensor-enhanced keystroke dynamics. In *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*, pages 92–111. Springer, 2014.

[16] Marian Harbach, Emanuel Von Zezschwitz, Andreas Fichtner, Alexander De Luca, and Matthew Smith. It's a hard lock life: A field study of smartphone (un) locking behavior and risk perception. In *Symposium on usable privacy and security (SOUPS)*, pages 213–230, 2014.

[17] Eiji Hayashi, Sauvik Das, Shahriyar Amini, Jason Hong, and Ian Oakley. Casa: context-aware scalable authentication. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, page 3. ACM, 2013.

[18] Eiji Hayashi, Oriana Riva, Karin Strauss, AJ Brush, and Stuart Schechter. Goldilocks and the two mobile devices: going beyond all-or-nothing access to a device's applications. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, page 2. ACM, 2012.

[19] Hassan Khan, Aaron Atwater, and Urs Hengartner. Itus: an implicit authentication framework for android. In *Proceedings of the 20th annual international conference on Mobile computing and networking*, pages 507–518. ACM, 2014.

[20] Hassan Khan, Urs Hengartner, and Daniel Vogel. Usability and security perceptions of implicit authentication: Convenient, secure, sometimes annoying. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*, pages 225–239, Ottawa, 2015. USENIX Association.

[21] Lingjun Li, Xinxin Zhao, and Guoliang Xue. Unobservable re-authentication for smartphones. In *NDSS*, volume 56, pages 57–59, 2013.

[22] Daniel C McFarlane. Comparison of four primary methods for coordinating the interruption of people in human-computer interaction. *Human-Computer Interaction*, 17(1):63–139, 2002.

[23] Nicholas Micallef, Mike Just, Lynne Baillie, Martin Halvey, and Hilmi Güneş Kayacik. Why aren't users using protection? investigating the usability of smartphone locking. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 284–294. ACM, 2015.

[24] Kenrick Mock, Bogdan Hoanca, Justin Weaver, and Mikal Milton. Real-time continuous iris recognition for authentication using an eye tracker. In *Proceedings of the 2012 ACM conference on Computer and communications security*, pages 1007–1009. ACM, 2012.

[25] Lawrence O'Gorman. Comparing passwords, tokens, and biometrics for user authentication. *Proceedings of the IEEE*, 91(12):2021–2040, 2003.

[26] Oriana Riva, Chuan Qin, Karin Strauss, and Dimitrios Lymberopoulos. Progressive authentication: Deciding when to authenticate on mobile phones. In *Proceedings of the 21st USENIX Conference on Security Symposium*, Security'12, pages 15–15, Berkeley, CA, USA, 2012. USENIX Association.

[27] Hataichanok Saevanee, Nathan L Clarke, and Steven M Furnell. Multi-modal behavioural biometric authentication for mobile devices. In *IFIP International Information Security Conference*, pages 465–474. Springer, 2012.

[28] Elaine Shi, Yuan Niu, Markus Jakobsson, and Richard Chow. Implicit authentication through learning user behavior. In *International Conference on Information Security*, pages 99–113. Springer, 2010.

[29] Hui Xu, Yangfan Zhou, and Michael R Lyu. Towards continuous and passive authentication via touch biometrics: An experimental study on smartphones. In *Symposium On Usable Privacy and Security, SOUPS*, volume 14, pages 187–198, 2014.

# Exploring Intentional Behaviour Modifications for Password Typing on Mobile Touchscreen Devices

Lukas Mecke[1,2†], Daniel Buschek[2†], Mathias Kiermeier[2‡], Sarah Prange[1,3,2†], Florian Alt[3]
[1]*University of Applied Sciences Munich, Munich, Germany, {firstname.lastname}@hm.edu*
[2]*LMU Munich, Munich, Germany, †{firstname.lastname}@ifi.lmu.de, ‡mathias.kiermeier@gmail.com*
[3]*Bundeswehr University Munich, Munich, Germany,{firstname.lastname}@unibw.de*

## Abstract

Behavioural biometric systems are based on the premise that human behaviour is hard to intentionally change and imitate. So far, changing input behaviour has been studied with the goal of supporting mimicry attacks. Going beyond attacks, this paper presents the first study on understanding users' ability to modify their typing behaviour when entering passwords on smartphones. In a prestudy (N=114), we developed visual text annotations to communicate modifications of typing behaviour (for example, gap between letters indicates how fast to move between keys). In a lab study (N=24), participants entered given passwords with such modification instructions on a smartphone in two sessions a week apart. Our results show that users successfully control and modify typing features (flight time, hold time, touch area, touch-to-key offset), yet certain combinations are challenging. We discuss implications for usability and security of mobile passwords, such as informing behavioural biometrics for password entry, and extending the password space through explicit modifications.

## 1 Introduction

The way we type on physical and on-screen keyboards is remarkably individual: Many studies have shown that people can be identified based on their typing rhythm [36], finger placement [11], and other such features of typing and touch behaviour [8, 37, 44]. This approach can be used, for example, to block unwanted access to technical systems, accounts, and personal mobile devices: Even if attackers gain knowledge of a password, they also have to enter it with the same behaviour as the legitimate user. The underlying assumption of such behavioural biometric authentication systems is that humans differ *implicitly* in how they type.

We present the first systematic exploration of a fundamentally different view: We study how users *explicitly* modify commonly utilised biometric features of their typing behaviour. Our goal in this paper is not to design a new authentication system but to better understand users' fundamental ability to control their typing behaviour. Better understanding such an ability to intentionally modify interaction behaviour is important in the light of a growing number of biometric security systems, as illustrated with the following use-cases:

*Extending the password space:* Instead of only using different characters to compose a password, each character could be entered in a different manner. For instance, although both use the same eight characters, "password" is different from "pass[*hold long*]word", where the user keeps the second "s" pressed for longer than her usual behaviour.

*Avoid leaking "natural" behaviour:* As more and more systems process behaviour, it might be a viable strategy for users to intentionally modify behaviour for some. For example, a user might authenticate on a work laptop using a modified typing rhythm when giving a presentation, to not reveal her "natural" typing behaviour, which she uses in (other) biometric systems, to a potential attacker. This strategy might also be used for authentication on the web or filling in a form in an unsafe environment, e.g., when using an unknown device.

*Recovering from a leak of behavioural data:* A leak of behavioural information implies that this biometric can no longer be used if we assume that behaviour is unchangeable. However, this is worth challenging. As an analogue example, some people decide to intentionally change the way they write their signature. Similarly, it might be possible to intentionally change, for example, password typing behaviour features to recover from a leak to be able to continue using this biometric.

In all these examples, users have reasons to intentionally modify aspects of their behaviour which they do not need to control for the underlying input method (e.g., typing rhythm does not matter for entering an email). Prior work on inten-


*USENIX Symposium on Usable Privacy and Security (SOUPS)* 2019.
August 11–13, 2019, Santa Clara, CA, USA.

tional changes of typing behaviour has exclusively studied this ability for attackers with technical support [4, 23, 24] or for limited features in desktop settings without changes and learning over time [14, 21, 33]. Thus, it still remains unclear to what extent users can control and modify fundamental biometric features of their mobile touch typing behaviour.

We address this gap by contributing: (1) Visual text annotations to communicate typing behaviour modifications, developed in a prestudy (N=114). (2) A lab study (N=24) using this scheme to investigate intentional modifications for different features and their combinations, for password typing on smartphones in two sessions a week apart. Based on the results, we discuss implications for mimicry attacks, research on behavioural biometrics, and usable passwords with intentional modifications.

The paper is structured as follows: After discussing related work (2), we develop a visualisation of typing behaviour (3), followed by our study design (4) and results (5) on intentional behaviour modifications. We conclude with a discussion (6).

## 2 Related Work

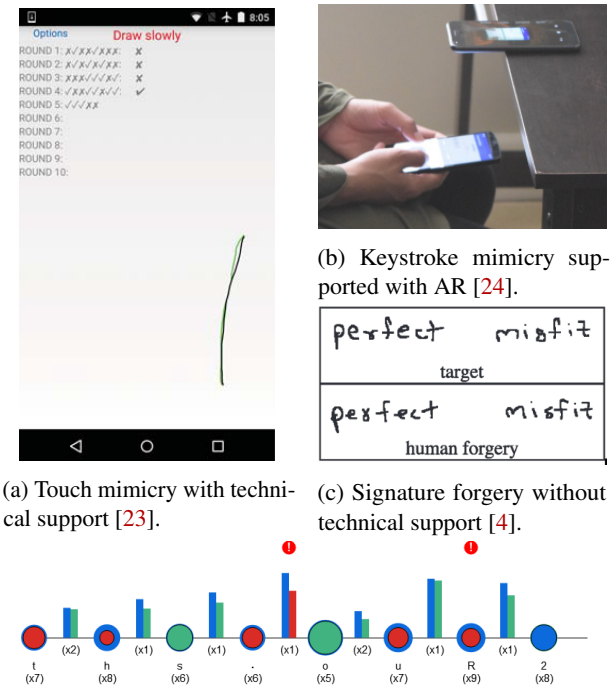
In this section, we relate our work to research on keystroke biometrics and mimicry attacks. These areas motivate our investigation of intentional modification of typing features and our choice of the specific features we studied.

### 2.1 Keystroke Biometrics

Our work is related to keystroke biometrics (or "keystroke dynamics"), which describe users' individual behavioural characteristics when entering text on a keyboard. This information can be used by the system to identify users, for example, to protect accounts, devices, and data. A rich body of related work examined this idea first for typing on physical desktop keyboards (for example, [29, 30]; survey [36]), then on early mobile phones with physical keys (for example, [7, 13, 15, 21, 22, 25, 46]). More recent work investigated keystroke biometrics for on-screen typing on smartphones (for example, [10, 11, 16, 44]; recent survey [37]), including keyboards operated via gestures instead of tapping [8].

For entering passwords in particular, recognising users based on *how* they enter the secret word provides an extra (implicit) layer of security [11], for example, to protect against cases in which the attacker got to know the password via shoulder surfing [32], smudge [2, 41] or thermal attacks [1].

Due to the origin of keystroke biometrics on physical desktop keyboards, the most commonly used typing behaviour features are temporal [36]: Users' typing is characterised by their typical *hold times* (i.e., time between key down and up event), and *flight times* (i.e., time between key up and down on the next key). Mobile touch devices offer further spatial features, such as touch area and offsets between touch locations and key centres. Offsets, in particular, showed higher




(a) Touch mimicry with technical support [23].

(b) Keystroke mimicry supported with AR [24].

(c) Signature forgery without technical support [4].

(d) Visualisation to support keystroke mimicry on a PC [38].

Figure 1: Several examples from related work for supporting mimicry attacks on (a) touch biometrics, (b, d) keystroke biometrics and (c) signatures. Images taken from cited papers.

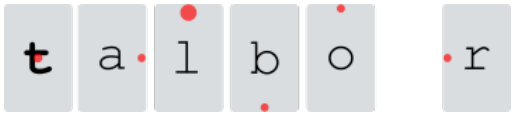
biometric value, that is, they facilitated more accurate distinction of users [10, 11]. Related work motivates our choice of features: hold time, flight time, offsets, and touch area.

In summary, related work on typing behavioural biometrics used features as they occur "naturally" as an *implicit* part of typing. Our work is fundamentally different: We examine these typing features as *explicit* and *actively controlled* by users, for example to increase the password space. In particular, we study how well users can indeed control these features when entering passwords on a smartphone.

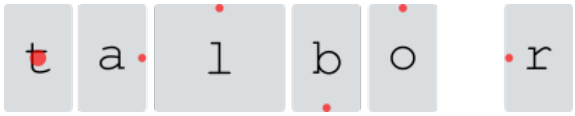
### 2.2 Mimicry Attacks

Attacks on keystroke biometric systems can be performed either automated or manually. Automated attacks use generative models to synthesise forgeries from observed data and were shown to be effective against handwritten signatures [4] and keystroke dynamics on a PC [28, 31, 34]. Some work also tested such attacks when proposing a new keystroke biometric system. For example, Stefan et al. found their system resistant against inputs generated from a first-level Markov model [35].

The most commonly considered attack on behavioural biometric systems is the so-called *mimicry* attack: Here, an impostor tries to manually reproduce (mimic) the (known) behaviour of a legitimate user to gain access.



(a) **'Bold Letter'** using bold font to indicate large touch area and circle size for hold time. Circle location shows offset, key gaps indicate flight time.



(b) **'Long Key'** using circle size for touch area and key width for hold time. Same as above: Circle location shows offset, key gaps indicate flight time.

Figure 2: Main design candidates for visualising target feature values for studying intentional behaviour modifications. Both were evaluated in our prestudy. Based on the results we decided to use the '*Long Key*' concept for our main study.

As a simple case, a *zero-effort attacker* model evaluates a biometric system against natural behaviour collected of other users who did not intend to actually bypass the system. While this model has been commonly used to evaluate vulnerability of behavioural biometric systems, related work found that it underestimates attack success [4, 31]. This calls for evaluations with means for more skilled and targeted attacks.

To support attackers in launching successful mimicry attacks they need to know the behaviour to imitate. In the case of handwritten text, for example, this could be a sample signature (cf. Figure 1–c). Researchers mounted successful mimicry attacks against touch input behaviour [23], keystroke dynamics on a PC [38], and keystroke dynamics on mobile phones [24].

Key to those attacks were systems which both visualise the target behaviour and provide the attacker with feedback on their attempts (cf. Figure 1). For example, Khan et al. [24] used augmented reality using a phone's camera to show visual cues on top of its view on another phone's keyboard. This guided correct timing and touch behaviour. In another approach they used audio stimuli to guide the timings.

In summary, prior work used representations and active modifications of typing behaviour to support mimicry attacks. In contrast, we aim to better understand the human ability to control mobile typing behaviour per se.

# 3 Prestudy: Developing a Visual Representation for Typing Behaviour Modifications

## 3.1 Selection of Features

There are a multitude of possible features that can be used for biometric authentication in the context of mobile touch interaction. An extensive list was compiled by related work [11] and covers 24 spatial, temporal and contact features. Khan et al. [24] found this extensive feature set hard to simultaneously control for their mimicry attack. They thus removed highly correlated features, resulting in a set of six: key hold time, flight time, down pressure, down area, down x, and down y.

We combine x and y together as touch offset. Furthermore, pressure and area were highly correlated on our test devices, since most Android phones[1] estimate pressure from area. We thus decided to omit pressure and used area directly.

To sum up, we decided to study a set of four features, namely *touch area*, *flight time*, *hold time* and *touch-to-key-offset* with the latter being two-dimensional (x, y).

## 3.2 Visualisation Design

We developed several designs that communicate modifications of the four features to instruct participants, for example, to perform a long key press for the second character in a password. We first tried simple markup (e.g., p– . **a̲s** . . sw̲–o̲**r** . d—) but found this representation to become cluttered quickly and to offer very limited expressiveness.

We thus chose a pictorial approach: We showed letters with a key metaphor to visualise behavioural changes (Figure 2). We explored a range of possible visual features, including offsetting the key or its label, writing bold or italic, and using underscores and coloured dots.

We narrowed the options down to two final designs (cf. Figure 2). Both used whitespace gaps between keys to indicate flight time and a red dot to indicate touch offset. One variant ('Bold Letter') visualised larger touch area by rendering the key in bold, and used the size of the offset dot to represent hold time. The other ('Long Key') used the size of the dot to visualise touch area, and key width to show longer hold time. While 'Bold Letter' resulted in a more compact format, 'Long Key' unified both temporal features on a shared axis (time flows from left to right). We conducted an online survey to determine our final design.

## 3.3 Online Survey

### 3.3.1 Survey Design and Procedure

To assess intuitiveness and readability of our designs, we created an online survey which showed example passwords with visualised modifications. Participants had to indicate which parts of the visualisation were used to encode which behavioural cues, without prior explanations. People did this for both designs in counterbalanced order. Afterwards, they were asked to rate on a 5-point Likert scale how intuitive and readable they found the two visualisations.

The survey was distributed over a university mailing list. It took 5 minutes to complete. Participants had a chance to win a €10 gift voucher.

[1] We used LG G6 phones in our study.

#### 3.3.2 Results

A total of 114 participants answered our survey (56 % female; mean age 27 years, range 18 to 63 years). Both *offset* and *flight time* were correctly interpreted by 90 % of the participants for both designs. *Area* and *hold time* were correctly interpreted by 81 % and 82 % in the 'Long Key' condition, respectively. However, these two features were only correctly interpreted by 50 % and 51 % in the 'Bold Letter' condition. 'Long Key' was rated as more intuitive (median=agree, median_bold=neutral) but 'Bold Letter' was rated to be more readable (median=strongly agree, median_long=agree). When asked for their preferred method, 59 % of the participants reported the 'Long Key' notation while 39 % voted for the 'Bold Letter' visualisation. The rest had no preference.

### 3.4 Final Visual Representation

We decided to use the 'Long Key' visualisation: It has the advantage of encoding temporal features on a shared axis and all features allow for continuous representation of values (in contrast to the binary bold letter).

In conclusion, we used the following visual encoding shown in Figure 2–b: *Touch-to-key-offset* is marked by a red dot at the position where the key should be touched. *Flight time* is represented by a whitespace gap between two key rectangles that scales with duration. Analogously, *hold time* is represented by scaling the width of the key rectangle with duration. Finally *touch area* is visualised by the size of the red dot used for offset (larger size indicates larger area).

## 4 Main Study

### 4.1 Study Design

As our study design is quite complex, the following subsections each explain one main component. The most complex one is *task*, which is given both as an overview and in detail.

#### 4.1.1 Passwords

In general, participants had to repeatedly enter given *passwords* ("football", "princess", "password"). While these three are obviously not great passwords in terms of security, we selected them since they have comparable properties and are common passwords[2]. Moreover, they do not require switching keyboard mode (e.g., between characters and symbols), which we wanted to avoid as a simplification for this first investigation into intentional typing behaviour modification. Similarly, we favoured simple passwords to ensure that task difficulty was mainly determined by behaviour variations and not affected by memorability or search time for rare symbols.

[2] https://www.teamsid.com/worst-passwords-2016/, last accessed 20.02.2019

#### 4.1.2 Features

We studied intentional modification of four features: *touch-to-key-offset* (on five levels: centre/left/right/top/bottom), *flight time* and *hold time* (both on two levels: default/long), as well as *touch area* (on two levels: default/large).

#### 4.1.3 Tasks

Participants solved 37 *tasks*, each using one of the three passwords. The tasks differed in various aspects, described below. While the design is complex, the overall goal was to cover six aspects, namely (1) different *passwords* with (2) different *feature modifications* at (3) different *locations* within each word. We also include (4) different *combinations* of features that are modified in the same password, either (5) at the same character/keypress (we call this *co-located*) or (6) *distributed* across several characters/keypresses within the word.

We iterated the task design several times by means of prestudy runs with two to three people in each version. We gradually narrowed the tasks down to an acceptable study duration of one hour. In full detail, the tasks used in the main study were structured and designed as follows (Figure 3):

*Natural tasks (1–3):* The first three tasks simply asked people to enter each password six times without presenting any intentional behaviour modifications.

*Modifying a single feature (tasks 4–15):* In each of these tasks participants had to modify one feature (e.g., hold time). There were three such tasks per feature, namely one per password (i.e., 4 features $\times$ 3 passwords = 12 tasks). Across the three tasks per feature, all feature levels occurred at least once, while covering different locations: The first task per feature modified the 2nd character of the password, the second task modified the 2nd and 7th characters, and the last task modified 2nd, 4th, and 7th characters. The assignment of passwords across these tasks was counter-balanced, such that modifications overall occurred in all passwords at all locations.

*Modifying two features (tasks 16–27):* In each of these 12 tasks people modified two features (for example, hold time *and* flight time). There were two tasks per combination of two features: The first had one modification on the 2nd character and the other on the 3rd (i.e., *distributed*). The second task had both modifications on the 7th character (i.e., *co-located*).

*Modifying three features (tasks 28–35):* In these eight tasks, participants had to modify three features, with two tasks per combination of three features: The first had modifications on the 2nd, 4th, and 7th character (*distributed*). The second one had all three modifications on the 5th character (*co-located*).

*Modifying four features (tasks 36 and 37):* Finally, participants had to modify four features: The first one had modifications on the 2nd, 4th, 6th, and 8th character (*distributed*), the last had all modifications on the 5th character (*co-located*).

The task order was not randomised, in favour of gradually increasing the number of modified features per password, which we suspected to have an influence on task difficulty.

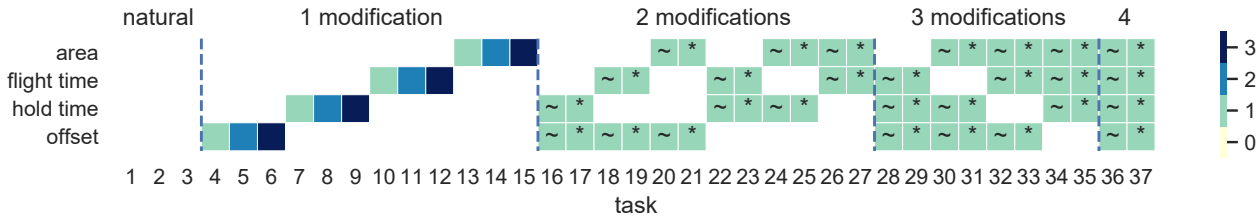



Figure 3: Overview over the tasks in each session. In the beginning (task 1–3) participants were asked to enter the passwords naturally, afterwards (task 4–15) a single feature had to be modified with increasing number of occurrences (colour of the cell). Thereafter, two (task 16–27), three (task 28–35) or four (tasks 36, 37) features had to be modified at once. All possible feature combinations were tested and features were either *distributed* (~) over the password or *co-located* (*) on a single key.

#### 4.1.4 Sessions

The whole procedure was repeated two times, in two sessions about a week apart. In this way, we observed the typing behaviour of each participant at two points in time.

#### 4.1.5 Summary

For the following report of our data analyses and results, it is useful to think of our study design as follows:

Tasks 1–3 are used to analyse natural (i.e., unmodified) behaviour, while the other tasks are used to analyse user behaviour when modifying the four behaviour features.

Note that from task 16 onward (i.e., all tasks with feature combinations), our study is a typical repeated measures design with: *number* of modifications (2, 3, 4) × *distributed* multiple modifications (distributed, co-located) × *session* (1st, 2nd). We use this for typical ANOVAs to study in particular the impact of modification of multiple features.

### 4.2 Apparatus

We developed an Android app that controlled the study process (e.g., counterbalancing, task progression, explanations).

The values used for scaling our visualisations (e.g., default flight time for default key gap) were informed by prestudy experiments and related work [10] (flight time 260 ms normal, 1000 ms long; hold time 80 ms normal, 300 ms long; area 0.2 normal, 0.4 large, unitless as reported by the Android API; offset x ±40 px, offset y ±70 px). To avoid visual clutter, we limited the scaling with minimum and maximum threshold values, beyond which the visualisation did not change.

We integrated a modified version of the Android open source project LatinIME[3] keyboard. This enabled us to log all typing events and touch features. To reduce distraction, we disabled the context menu for special characters shown on long press. In addition, our study app logged the expected key and behaviour modifications, as well as the current user and task for each keystroke.

[3] https://android.googlesource.com/platform/packages/inputmethods/LatinIME/, last accessed: 22.02.2019

### 4.3 Procedure

Upon arrival, participants were introduced to the goal of the study and were asked to sign a consent form to permit use of the collected data. After an initial demographics questionnaire they performed the tasks (cf. Figure 3) as described in section 4.1.3 on our test device. We asked participants to enter passwords with their right thumb to keep results comparable.

When first confronted with a new type of modification, participants got a short explanation of what to do and prior to every task they had the option to train entering the password. Except for the tasks without modifications (natural tasks) they were provided with real time feedback, using our visualisation, to show their behaviour next to the expected one. Every task had to be completed successfully six times and without feedback. The number of attempts was not limited.

Each task was followed by a short Likert questionnaire containing the statements: (1) *"I was able to adjust to the specified behaviour."*, (2) *"I was successful in completing the task."*, and (3) *"The task was difficult for me."*.

After completing all tasks, participants were asked to come up with a modified password on their own and could take notes to remember it. The same process was repeated in the second session, excluding the initial demographics questionnaire. Creating a custom password was replaced with recalling and performing the password from the previous session. After the second session we conducted a short interview. Sessions were scheduled one week apart.

### 4.4 Participants

Study invitations were distributed over a mailing list of our local university. Requirements were right-handedness and familiarity with typing on mobile phones. We recruited a total of 24 participants (14 female; mean age 27 years, range 14 to 54 years). Half of participants were in their twenties. 58 % were students, 30 % were employed, and the remaining ones were in school. Participants were compensated with €20 for completing the whole study.

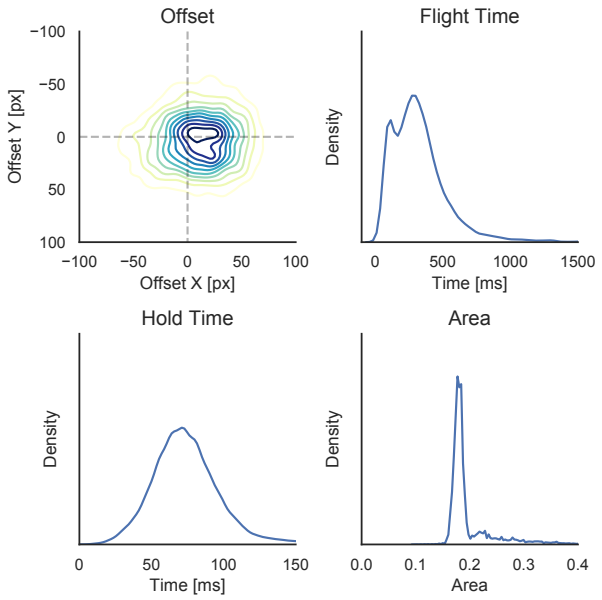



Figure 4: Overview of participants' natural typing behaviour (i.e., typing without being presented with any modifications), as measured in the first three tasks of each session.

## 5 Results

Significance tests were conducted using ANOVA with Greenhouse-Geisser correction and Bonferoni corrected post-hoc tests (significance at alpha level $p < 0.05$). If not reported otherwise, data for analyses is aggregated for both sessions.

As a first overview, we report key descriptive measures: The grand mean task completion time across all tasks(i.e., completing all six successful password entries of a task) and participants was 38.3 seconds . For typing speed, the grand mean was 28.7 words per minute (WPM [43]). The grand mean of the number of incorrect entries per task was 1.74.

We report on participants' natural typing behaviour (5.1), their ability to modify it (5.2), and their accuracy in doing so (5.3). We analyse the effect of multiple simultaneous modifications (5.4) and the impact of modifications on individuality of behaviour (5.5). We conclude with details on technically detecting modifications (5.6) and participant feedback (5.7).

### 5.1 Natural Behaviour

We first report on "natural" behaviour – typing *without* any modification instructions (tasks 1–3). Figure 4 presents the results. They match our expectations based on related work:

Touch offsets are slightly shifted to the lower right, as typical for input with the right thumb [9]. Moreover, median flight time (290 ms) and hold time (72 ms) are in line with related work [10] and close to the ones we chose as defaults for scaling key width and gaps in our visualisation (flight time 260 ms, hold time 80 ms). Thus, our chosen values indeed matched people's natural behaviour.

| Feature | Measure | target | session | target * session |
|---|---|---|---|---|
| Offset | absolute x | $.777^a$ | | |
| | absolute y | $.890^a$ | | $.015^c$ |
| | relative (error) | $.082^b$ | | |
| Flight time | absolute | $.785^a$ | | $.010^c$ |
| | relative (error) | $.332^a$ | $.038^b$ | |
| Hold time | absolute | $.848^a$ | | |
| | relative (error) | $.624^a$ | | |
| Touch area | absolute | $.737^a$ | | |
| | relative (error) | $.930^a$ | | |

*a*: $p < .001$, *b*: $p < .005$, *c*: $p < .05$, empty cells not significant

Table 1: ANOVA results for ability (1) to modify behaviour (absolute, Section 5.2) and (2) to replicate target feature values (relative i.e., error, Section 5.3). The last three columns show the effect sizes ($\omega^2$) for *target* value (i.e., the feature value communicated via our text annotation), *session*, and their interaction. See text for results from post-hoc tests.

Touch area significantly correlated with x location of the target key (r=-0.252, p<.001): Due to thumb stretching, typing keys on the left of the keyboard resulted in a flatter thumb posture and thus larger touch area. Flight time showed a main and secondary peak (Figure 4). The latter was caused by zero finger travel distance for "double letters" (e.g., pa*ss*word).

### 5.2 Ability to Modify Behaviour

Figures 5 and 6 visualise the distribution of the behavioural features for different *target values*, i.e., expected feature values shown by our visualisation. Next, we report on statistical tests comparing these distributions per feature (see Table 1). Here we report on the post-hoc tests and further details:

For all features, post-hoc tests showed that directions of differences were as expected (e.g., offset significantly further to the left for *left*, flight time significantly longer for *long*).

For vertical offset and flight time, the interactions of session and target were significant (see Table 1), yet the small effect sizes and visual inspection of descriptive plots indicated that this was too tiny to warrant meaningful interpretation.

In summary, the significant results of these statistical tests confirm the "big picture" visible in Figure 5 and Figure 6: For all features, people significantly modified their behaviour in the direction indicated by our visualisation.

### 5.3 Ability to Replicate Target Feature Values

The previous section investigated differences in absolute feature values. It is also interesting to analyse how *accurately* people were able to replicate modifications. To this end, Figure 7 visualises the distribution of participants' errors when reproducing the target values indicated by our visualisation for each feature. Table 1 summarises the ANOVA results.

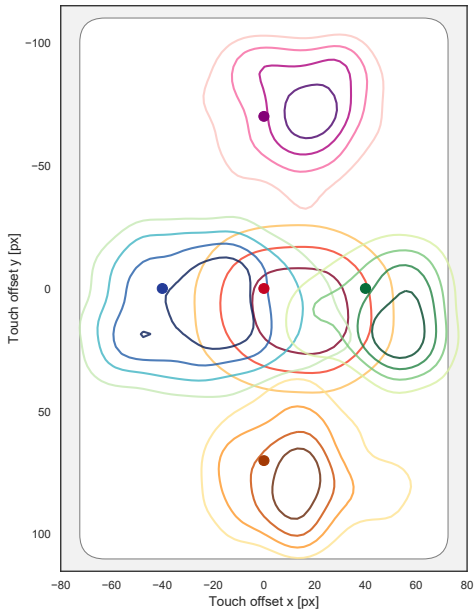



Figure 5: Overview of users' modified touch-to-key offsets: Provoking offset modifications resulted in clear differences in thumb placement. The rectangle indicates key borders.

For *offset*, post-hoc tests revealed errors to be significantly smaller for the target *right* compared to *left* (p=.010, d=-.773), *top* (p=.008, d=-.783), *bottom* (p=.011, d=-.765) and *default* offset (p=.027, d=-.685).

For *flight time*, we found errors to be significantly smaller for the *default* time than the *long* one (p<.001, d=-1.488), as well as for observations from the *second* session compared to the *first* (p=.004, d=-.645). The latter matches the observation that people typed slightly faster in the second session.

Regarding *hold time*, post-hoc tests showed errors to be significantly smaller for the *default* time compared to the *long* one (p<.001, d=-1.844). Finally, for *touch area*, we found errors to be significantly smaller for the *default* area size compared to the *large* one (p<.001, d=-4.470).

In summary, these results confirm that participants significantly modified their behaviour, namely towards the values indicated by our visualisation. In addition, people are more accurate in producing the default feature values compared to the more extreme ones, likely because the latter are further away from "natural" typing behaviour.

## 5.4 Impact of Modifying Multiple Features

Here we report on users' ability to modify multiple features in one password. Table 2 summarises the ANOVA results. Post-hoc tests and further details follow below.

### 5.4.1 Impact on Time, Speed, and Incorrect Entries

For *task completion time*, post-hoc tests revealed that three modifications resulted in significantly longer times compared to two (mean 40.70 s vs 36.36 s; p<.005, d=0.543); descriptively, this was also true for four modifications compared to two, yet not significantly so (p=.064). Moreover, distributed multiple modifications took significantly longer than co-located ones (mean 42.33 s vs 34.33 s; p<.01, d=1.397). People were also significantly slower in the first session than in the second one (mean 39.76 s vs 36.90 s; p<.05, d=0.444).

For *typing speed*, all pairwise comparisons of number of modifications were significant (all p<.001), with slower typing for higher numbers (mean 2: 30.18 WPM, 3: 27.33 WPM, 4: 25.15 WPM). Moreover, distributed multiple modifications were typed significantly slower compared to co-located ones (mean 26.91 WPM vs 30.46 WPM; p<.001, d=-2.445).

Finally, significantly more *incorrect password entries* occurred for distributed compared to co-located multiple feature modifications (mean 2.44 vs 1.45; p<.005, d=0.677).

These results show that users take significantly longer to enter passwords as the number of modified features increases, in particular if behaviour is modified for multiple features across different characters (i.e., *distributed*). In that case, people also produce significantly more incorrect password entries.

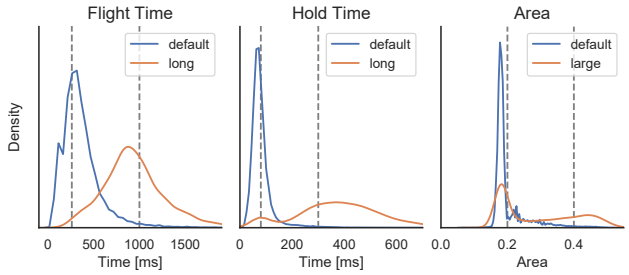



Figure 6: Overview of participants' modified typing behaviour across both sessions. Overall, this figure shows that presenting modifications via our visualisation provoked clear differences in the typing features (flight time, hold time, area; for touch offset see Figure 5). Vertical lines indicate the target values.

| Measure | number of mod. | session | distributed | number * distributed | session * distributed |
|---|---|---|---|---|---|
| Offset error | | | | | |
| Flight time error | .93$^a$ | .017$^b$ | .109$^a$ | .023$^c$ | |
| Hold time error | .166$^a$ | | .178$^a$ | | |
| Touch area error | .039$^a$ | | | .018$^a$ | |
| Task compl. time | .032$^c$ | .015$^c$ | .224$^a$ | .039$^c$ | |
| Typing speed | .172$^a$ | | .232$^a$ | .079$^a$ | .002$^c$ |
| Incorrect entries | | | .114$^b$ | | |

*a*: p < .001, *b*: p < .005, *c*: p < .05, Empty cells not significant.

Table 2: Overview of ANOVA results for the impact of modifying multiple features on performance measures (Section 5.4.1) and ability to replicate target feature values (i.e., error, Section 5.4.2). Columns show effect sizes ($\omega^2$) for *number* of modifications, *session*, and *distributed* multiple feature modification, plus interactions. See text for details.

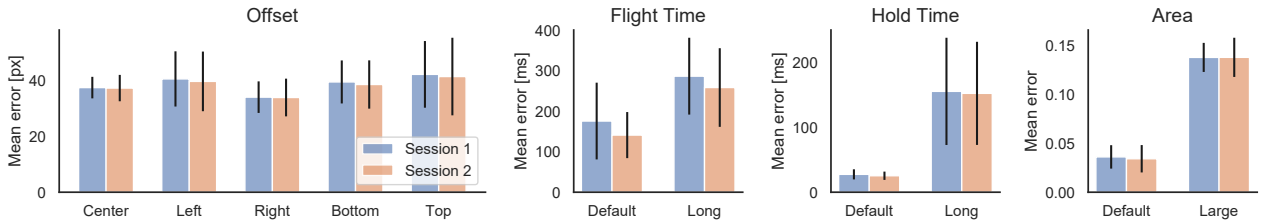



Figure 7: Observed derivation of participants' behaviour from the target values of the given modifications for both sessions. Participants were generally better at reaching the target value for the default level. For offsets, lowest error occurred for touches to the right, since this coincides with natural thumb offset [9]. In contrast to the other features, for flight time accuracy increased from the first to the second session, indicating a learning effect.

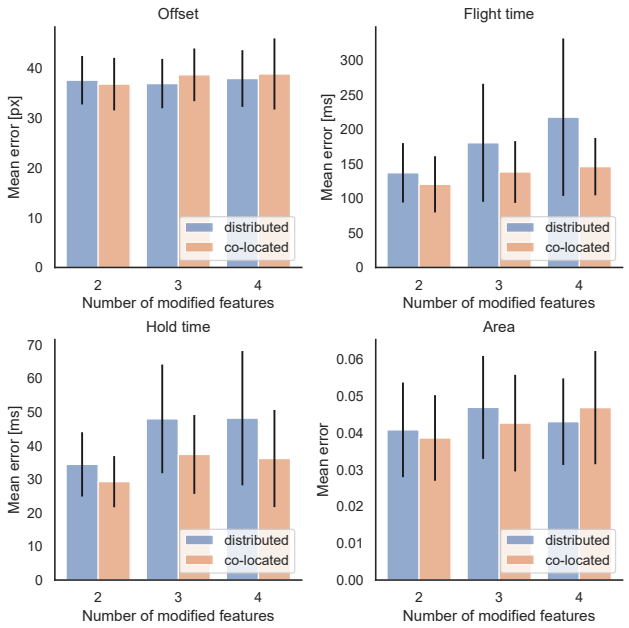



Figure 8: Participants' ability to replicate given behaviour depending on the number of features that had to be modified in one password and whether those features were co-located on a single key or distributed over the password.

### 5.4.2 Impact on Replicating Target Feature Values

Figure 8 shows participants' behaviour deviation from the given target behaviour (i.e., error), based on the *number* of features that had to be controlled within a single password and whether those features were *co-located* or *distributed*.

For *offset*, we found no significant effects (cf. very stable distribution of errors in Figure 8).

For *flight time*, errors were significantly lower for *co-located* modifications compared to *distributed* ones (p<.001, d=-.965), and for the second session compared to the first one (p=0.02, d=-.699). Regarding the number of modified features we observed significantly lower errors for two compared to three (p<.001, d=-1.149) and four (p<.001, d=-1.522), as well as for three compared to four modifications (p<.001, d=-.867).

Post-hoc tests for *hold time* revealed significantly lower errors for *co-located* features (p<.001, d=-1.004) and for two modified features compared to both three (p<.001, -1.479) and four (p<.001, d=-1.073) modifications.

Finally, for *touch area*, post-hoc tests showed significantly lower errors for two modified features compared to both three (p<.001, -1.565) and four (p<.001, d=-0.868) modifications.

Results are in line with the findings from the previous section. Participants generally performed better when features were *co-located* (i.e., not distributed over the password, Figure 8) and performance decreased for increasing *number* of modifications. Offset error was stable regarding all factors.

### 5.4.3 Impact on Subjective Rating

Participants answered three Likert items after each task: (1) *"I was able to adjust to the specified behaviour."*, (2) *"I was successful in completing the task."*, and (3) *"The task was difficult for me."* We compared users' ratings on these questions between tasks with co-located and distributed modifications: Wilcoxon signed-rank tests revealed significant differences for all three questions (Q1: Z=3.828, Q2: Z=4.074, Q3: Z= -3.765, all p<.001). Thus, participants subjectively perceived tasks with multiple feature modifications at the same character as significantly easier (i.e., better able to adjust behaviour, higher success, less difficult), compared to tasks with feature modifications distributed over several characters.

## 5.5 Impact of Modifications on Individuality

The previous analyses have shown behaviour differences *within users*, caused by modification instructions. Complementary, we now investigate how natural behaviour differences *between users* are influenced by modifications. This is interesting, for example, to inform behavioural biometric security layers. We will return to this in our discussion.

We thus compared the individuality (or "biometric value" [10]) of typing behaviour between natural and modified behaviour. To do so, we employed a user identification model [10, 12]. Note, that we do *not* intend to present this model as a practical biometric identification system. We rather

use it as an *analysis tool* to quantify the impact of explicit behaviour modifications on individuality. Thus, we are not interested in optimising identification accuracy, but in measuring the differences obtained on natural and modified behaviour.

### 5.5.1 Evaluation Scheme

We used the established Gaussian model for mobile touch typing, with a Gaussian distribution per feature per key [3, 19, 20, 45]. For touch location, for example, it defines the user's spread of touch points when aiming for that key. Thus, each user $u$ is represented by a set of Gaussians (the model $m_u$), fitted to the touches from the training set for that user. We used the data from the first session to fit these models.

For each user $u$, we then fed the data from $u$'s second session to this user's model $m_u$, which yields likelihoods for $u$ (for an ideal model, these should be high). In particular, we computed the joint likelihood for all touches for each task $t$, that is, the likelihood that $u$ is the one who typed the password in task $t$. Note that the features are per touch, not per password. Complementary, we fed the data from all other users $v \in U \backslash \{u\}$ to the model $m_u$ as well (for an ideal model, these likelihoods should be lower). We repeated this for all pairs of users $u, v \in U$, such that we obtain 24 (user models) $\times$ 24 (user data) likelihoods per task. We repeated the whole analysis twice, once for natural and modified typing data.

On these likelihoods, we computed the standard measures for typing biometrics (e.g., see [10, 36]): receiver-operating-characteric (ROC) curve, area-under-curve (AUC), and equal error rate (EER). An EER of X% means that in X% of password entries the legitimate user would be incorrectly rejected while also X% of attacks would pass unnoticed.

### 5.5.2 ROC Analysis Results

Figure 9 shows ROC, AUC and EER. Compared to random guessing (dotted line, 0.5 AUC), both natural and modified typing clearly yield biometric information. The values are in line with related work using this model for password typing on smartphones with the right thumb in the lab [11]. The results also show that people retain aspects of their individual behaviour when asked to perform the same modifications.

The key observation is the *gap* between the curves in Figure 9. It quantifies the loss in individuality: To summarise, when measured using an established typing model, individuality of participants' typing behaviour was *reduced* by intentional behaviour modifications such that AUC dropped by .07 (relative -8.9 %) and EER increased by .06 (relative +20.7 %).

## 5.6 Detecting Modifications

Finally, we analysed how well behaviour modifications can be technically detected. This is important, for example, to build an authentication system that allows these modifications to be used as part of a password. For instance, to check a password like "pass[*hold long*]word", the system needs to be able to distinguish between normal and long hold times.

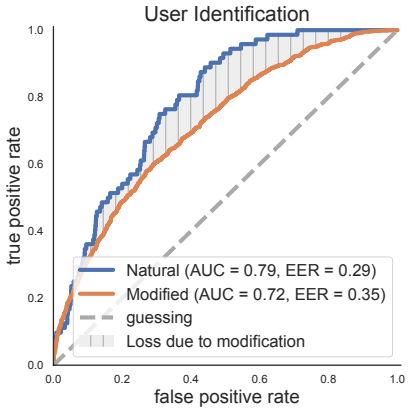



Figure 9: Impact of behaviour modification on individuality of typing behaviour, quantified by measuring the difference (shaded area) between ROC curves for user identification on natural (blue) vs modified (orange) typing. Typing behaviour becomes less individual through performing modifications. Clear individual characteristics remain, as evident from the modified (orange) line well above chance (dashed line).

We employed Random Forest classifiers with 100 trees and default parameters[4]. We used all typing features as input (hold time, flight time, area, offset x, y) and trained one model per modification (e.g., to classify normal vs long hold times).

We used leave-one-user-out evaluation across sessions: For each user $u$, we trained the classifier on the first session's data of all users except for $u$. We tested this model on $u$'s data from session two. Thus, the model could be shipped pre-trained and would not require data collection during enrolment.

We report mean (std) classification accuracy over all users: hold time 97.9 % (1.36 %), flight time 96.14 % (1.84 %), area 94.71 % (1.16 %), and offset 94.29 % (0.96 %). Note, that the remaining error includes user errors (e.g., user accidentally performed normal instead of long hold time). For these user errors, the model has to give an incorrect classification.

These results demonstrate that modifications can be reliably detected. It is thus technically feasible to implement an authentication system that allows users to use these modifications as part of their password. We provide the model code and trained model as part of the material for this paper (see Section 8) to facilitate implementations and further research on such password systems.

## 5.7 User Feedback

After the study we conducted short interviews: Half of the participants (12) stated to be interested in using passwords

[4] https://scikit-learn.org/stable/modules/ensemble.html#forest, *last accessed 20.02.2019*

with behavioural modifications and four were strictly against it. The other eight had concerns (e.g., security, being able to reproduce their behaviour under different circumstances or technical feasibility of such a system), but stated they would be interested in using a system utilising intentional modifications if those concerns could be addressed.

Many participants said they struggled with offset modifications as they would often hit the wrong key. Some also had difficulties distinguishing large area and long hold time.

When creating passwords, users often first observed their natural behaviour to then emphasise it. For example, P20 stated: "*When I created the password I first typed it and observed what I automatically did. For example I typed a 'g' rather to the left, entered a 'b' rather [long]; That's what I adjusted [the password] to.*". Another common strategy was putting modifications at salient positions, such as at the beginning of words or syllables.

# 6 Discussion

## 6.1 Controlling Password Typing Behaviour

As a key insight, we revealed that people are able to significantly modify temporal and spatial features of their mobile typing behaviour in given directions. It is also possible to train a model that distinguishes between these features levels (e.g., default vs long press) with high accuracy (Section 5.6).

People were more accurate (i.e., deviated less from target feature values) in reproducing default values rather than extreme ones. We thus conclude that people are better at replicating behaviour that is close to their natural behaviour.

For flight time, accuracy was higher in the second week. We attribute this to people getting accustomed to our device, modifications, and tasks, indicating a learning effect.

In some cases participants performed default behaviour when expected to show a modification (see secondary peaks in distributions in Figure 6), likely due to the cognitive load of actively controlling their actions, especially when modifying multiple features. Controlling touch area is partly affected by the usage of the right thumb, which naturally leads to larger areas towards the left of the screen, due to stretching.

## 6.2 Modifying Multiple Behaviour Features

Overall, modifying an increasing number of behaviour features in a password becomes significantly more difficult to control. A possible explanation is the likely higher cognitive demand for intentionally modifying several aspects of typing behaviour, as supported by participants' comments and a higher number of incorrect inputs.

Specifically, modifying multiple features at different characters within one password ("distributed modification") is significantly more difficult than modifying multiple features at the same character ("co-located modification"). This conclusion is supported by all quantitative measures (task completion time, typing speed, incorrect entries, error measures), as well as participants' subjective Likert ratings.

Control of temporal features particularly suffers when other modifications are present, likely since focusing on those others distracts users from keeping the timing for the temporal modifications. Controlling spatial features is more robust.

In summary, our findings show that multiple features are harder to control when spread over multiple different characters; in particular, if temporal modifications are involved.

## 6.3 Methodology

We developed a visual text annotation scheme (Figure 2) to communicate target behaviour modifications. We chose this approach to be able to use text entry research's most common and established transcription task (i.e., enter given text) with our new concept of intentional behaviour modifications.

An alternative would have been to visualise desired feature values directly on the keyboard (e.g., show cross-hair on the key for offset modifications). However, this would have turned the task into a *reaction* exercise (i.e., hitting such cross-hairs), which likely leads to different behaviour. This approach also borrows heavily from the technical support work on mimicry attacks. Yet we were interested in users' ability to modify behaviour without such scaffolding. With our task, we thus gave clear instructions while participants were left to implement those modifications as they saw fit.

Future work could compare the two approaches. For example, work on systems for mimicry attacks could use our results here as a baseline for unsupported modification ability.

## 6.4 Deployment

As shown in Section 5.6, it is possible to reliably detect behaviour modifications, which enables building authentication systems that utilise them as part of a password. With backends that store passwords as hashes of strings, this could be easily integrated by inserting a special symbol depending on the preceding character's modification (e.g., "pass$holdlong$word" where $ stands for any character not allowed to be used directly for passwords in the system). Therefore, this technique can potentially be used in any context that passwords are currently used in – given that client software and hardware are capable of detecting modifications. For non-touch keyboards, only temporal features would be available.

Moreover, our visualisation (Section 3) could give users feedback on their typing, analogous to revealing entered characters in a password field on demand.

Finally, it is not clear how different devices and keyboard layouts influence behaviour and control, which could be investigated in future work.

## 6.5 Implications for Usable Passwords with Intentional Behaviour Modifications

Intentional behaviour modifications increase the space of possible passwords. We focused on the fundamental ability of users to control behaviour features. Our results offer plenty of opportunities for future work, e.g., investigating observability and memorability.We summarise practical recommendations for usable passwords with behaviour modifications:

Flight time, hold time, and touch-to-key offset present suitable behaviour features for intentional modification for password typing on smartphones. Modifications of touch area for thumb input should be avoided. Area is harder to control since it is partly determined by stretching of the thumb.

Flight time and hold time can be controlled on two levels (normal vs long). Offsets can be controlled on five levels though they were the most difficult modification for participants. We see several options to improve this for future work. This includes tolerance for miss-typing (i.e., accepting input that hits a neighbouring key in the direction of the executed modification) and using offset modifications only with larger keys (e.g., on tablets or for PINs). Modifying offsets may also be easier when typing with a different finger that allows for more precision (e.g., index). Modifying behaviour for one character in multiple ways should be favoured over distributing feature modifications across several characters. Combinations of feature modifications across multiple characters in particular for temporal modifications should be avoided.

Based on user feedback after creating own passwords, a promising creation strategy is to observe one's own natural behaviour and add emphasising modifications.

## 6.6 Implications for Mimicry Attacks

Related work [10, 11] found that spatial features (particularly offsets) have higher biometric value, that is, they lead to more accurate user identification, compared to temporal features. Our results show that it is difficult to intentionally modify multiple temporal features, or temporal features combined with others. In contrast, for modifying offsets, users are not inherently under time-pressure when controlling them.

We thus revealed a novel trade-off: Spatial features have higher biometric value than temporal ones in the literature, yet they might be easier for informed attackers to modify. Future work can investigate such mimicry attacks: In particular, our results suggest 1) to compare mimicry attacks on biometric systems that use either spatial or temporal features; and 2) to compare such attacks for "victims" that do or do not intentionally control these features as part of their passwords.

In contrast to most previous work on mimicry attacks, these new study ideas do not focus on technical support for attackers or specific protection methods, but rather on better understanding the fundamental human capabilities for copying and controlling otherwise uncontrolled input behaviour details.

## 6.7 Implications for Biometrics Research

We showed for the first time that when multiple people follow the *same* modification instructions, their mobile typing behaviour becomes less distinguishable (here relative +20.7 % equal error rate for user identification across sessions).

Earlier work on typing on desktop keyboards [14, 33] and phones with physical key pads [21] discussed "artificial rhythms" (e.g., inserting a pause), which *increased* biometric value, contradicting our results. This difference may be due to typing on touchscreens in our work and the fact that related work studied behaviour in one session only, ignoring changes over time. Moreover, users received "open" instructions to modify the rhythm as they liked and thus likely responded in more individual ways [33]. Typing biometrics for desktops can only utilise temporal features. In contrast, mobile touchscreens enable rich spatial features and it can be difficult to coordinate modifications of multiple features in one password entry. This might have caused less consistent behaviour across sessions, reducing accuracy of user identification.

On one hand, this suggests that authentication systems need to be careful with applying *both* behavioural biometrics (e.g., as an extra security layer) and intentional modifications (e.g., for extended password space). On the other hand, suggesting *different* modifications to different users could improve biometric value, as we find users able to follow modifications of the most important features in typing biometrics.

Other work examined related ideas that might be investigated in our context as well: (1) nudging users towards creating more diverse lock patterns via subtle visual cues [40]; and (2) facilitating user exploration of "original" behaviour [42].

Our results guide future work on the idea of provoking more diverse behaviour: For example, a future study could ask users to set up a password not only with composition instructions (e.g., minimum length), but also suggest (random) behaviour modifications for how to enter it. Based on our results, we expect to achieve higher biometric value in this way, compared to 1) suggesting no behaviour modifications, or 2) suggesting the same modification to all users.

## 6.8 Security Considerations

Using intentional behaviour modifications impacts password capture and guessing attacks [6]. *Capture attacks* like smudge attacks [2] may be deflected, as temporal features leave no marks. Video-based attacks like shoulder surfing [32] or thermal attacks [1] may still be possible, though potentially harder, as extracting exact timings may prove difficult and fingers occlude the concrete touch points as long as no feedback is given (compare 6.4). Phishing may only be successful if the interface can capture and transmit modifications.

Assuming random passwords and modifications, adding modifications makes both online and offline *guessing attacks* harder (Table 3). Including one modification adds up to about

| password length | 8 | 7 | 6 | 5 |
|---|---|---|---|---|
| no modifications | **49.36** | 43.19 | 37.02 | 30.85 |
| 1 modification | **55.14** | 48.77 | 42.38 | 35.94 |
| 2 modifications | 59.84 | **53.27** | 46.63 | 39.90 |
| 3 modifications | 63.90 | 57.10 | **50.20** | 43.16 |

Table 3: Entropy (*bits*) of random passwords with and without (random) modifications on an alphabet of 72 characters (upper and lower case letters, numbers and 10 special characters).

5 bits of entropy (calculations in Appendix A). Thus, modifications may enable shorter passwords maintaining similar entropy. For instance, under the given assumptions, an eight character password can be reduced to six characters when using exactly 3 modifications. This is promising as passwords on mobile devices tend to be weaker and harder to enter [27].

Notice that these are upper bounds; there may be common patterns of choosing modifications, which reduce theoretical entropy in practice (e.g., participants reported to choose beginnings of words or syllables for modifications, cf. Section 5.7). Moreover, focusing modifications on a single key instead of spreading them out makes guessing easier. However, our calculations assume that the attacker knows the exact number of modifications, thus (slightly) underestimating entropy. While suggesting concrete modifications might solve some of those drawbacks it may introduce usability issues. We suggest practical security as an area for future work.

### 6.9 Limitations

We examined a limited set of typing features with a commonly used keyboard app (modified Google open source keyboard). We did not measure pressure or shape features from the full capacitive image (cf. [26]). Nevertheless, we covered the most commonly used temporal and spatial typing biometrics features (cf. [36, 37]), found to be the most important ones among a larger set for mobile password typing [11].

To avoid an impact of password complexity we chose a limited set of easy passwords for our study. Our findings may not generalise to more complex passwords.

To keep an acceptable study duration, we only observed one-handed use with the right thumb. This is one of the most considered postures in research [5, 17, 18, 45] and one of the most frequently used ones in daily life [10]. All participants were right-handed and used to this posture. Future studies could compare our results to typing with the index finger.

During analysis of the results we noticed that the target behaviour in task 34 contained an additional hold time modification instead of the intended flight time modification. Thus the combination of area, hold and flight time was not tested.

Our sample is biased towards younger people and might not represent the overall population. Finger precision and timing might change with age (cf. [39]). Future work could compare our results to samples with children and older adults.

## 7 Conclusion

Typing behaviour can be analysed to identify users based on features such as typing rhythm [36] and finger placement [11]. So far, research had studied these features as they occur "naturally" as an implicit, uncontrolled part of typing, or in the context of supporting mimicry attacks with technical means.

This paper addresses the gap in the literature with the first study on users' ability to intentionally modify their behaviour when typing passwords on smartphones: We developed a novel visual text annotation in a prestudy (N=114), before using it to study intentional modifications in the lab (N=24).

Overall, our results reveal that users can successfully modify the features most commonly used in typing biometrics systems for smartphones. This fundamental insight has several implications for users, threat models, and biometrics research. We conclude by outlining some of them here:

It is worth investigating further the idea of using intentional modifications as a part of passwords. This could extend the password space (e.g., "password" vs "pass[*hold long*]word") and possibly also reduce observability, as attackers would have to guess the modification, not just the entered word.

Our results also motivate novel research directions for touch and typing biometrics systems: These might suffer from "standardizing" typing behaviour across users with given modifications, as revealed in our study. However, nudging different users to use different modifications in turn promises to increase user identification accuracy (cf. [40]).

Related, threat models for evaluating such biometric systems need to take into account that some target behaviours are inherently more difficult to attack: In particular, our results strongly motivate comparing attacks that require modifying temporal vs spatial features to mimic the victim's behaviour.

Overall, we show the rich capabilities of users to intentionally control typical input behaviour features previously considered as an implicit "information byproduct" of interaction. With this work, we hope to spark new research and discussion regarding the use of behaviour-aware security systems that go beyond the view of a passively analysed user to take into account these human capabilities.

## 8 Project Resources

Material for this paper is available at: https://www.unibw.de/usable-security-and-privacy/downloads/datasets/intentional-behaviour-modifications


## Acknowledgements

Work on this project was partially funded by the Bavarian State Ministry of Education, Science and the Arts in the framework of the Centre Digitisation.Bavaria (ZD.B). This research was supported by the Deutsche Forschungsgemeinschaft (DFG), Grant No.: AL 1899/2-1.

## References


[1] Yomna Abdelrahman, Mohamed Khamis, Stefan Schneegass, and Florian Alt. Stay cool! understanding thermal attacks on mobile-based user authentication. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 3751–3763. ACM, 2017.

[2] Adam J. Aviv, Katherine Gibson, Evan Mossop, Matt Blaze, and Jonathan M. Smith. Smudge attacks on smartphone touch screens. In *Proceedings of the 4th USENIX Conference on Offensive Technologies*, WOOT'10, pages 1–7, Berkeley, CA, USA, 2010. USENIX Association.

[3] Tyler Baldwin and Joyce Chai. Towards online adaptation and personalization of key-target resizing for mobile devices. In *Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces*, IUI '12, pages 11–20, New York, NY, USA, 2012. ACM.

[4] Lucas Ballard, Daniel Lopresti, and Fabian Monrose. Forgery quality and its implications for behavioral biometric security. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 37(5):1107–1118, 2007.

[5] Joanna Bergstrom-Lehtovirta and Antti Oulasvirta. Modeling the functional area of the thumb on mobile touchscreen surfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 1991–2000, New York, NY, USA, 2014. ACM.

[6] Robert Biddle, Sonia Chiasson, and Paul C Van Oorschot. Graphical passwords: Learning from the first twelve years. *ACM Computing Surveys (CSUR)*, 44(4):19, 2012.

[7] A. Buchoux and N. L. Clarke. Deployment of Keystroke Analysis on a Smartphone. In *Australian Information Security Management Conference*, 2008.

[8] Ulrich Burgbacher and Klaus Hinrichs. An implicit author verification system for text messages based on gesture typing biometrics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 2951–2954, New York, NY, USA, 2014. ACM.

[9] Daniel Buschek and Florian Alt. TouchML: A machine learning toolkit for modelling spatial touch targeting behaviour. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*, IUI '15, New York, NY, USA, 2015. ACM.

[10] Daniel Buschek, Benjamin Bisinger, and Florian Alt. ResearchIME: A mobile keyboard application for studying free typing behaviour in the wild. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 255:1–255:14, New York, NY, USA, 2018. ACM.

[11] Daniel Buschek, Alexander De Luca, and Florian Alt. Improving accuracy, applicability and usability of keystroke biometrics on mobile touchscreen devices. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 1393–1402, New York, NY, USA, 2015. ACM.

[12] Daniel Buschek, Alexander De Luca, and Florian Alt. Evaluating the influence of targets and hand postures on touch-based behavioural biometrics. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 1349–1361, New York, NY, USA, 2016. ACM.

[13] P Campisi, E Maiorana, M Lo Bosco, and A Neri. User authentication using keystroke dynamics for cellular phones. *IET Signal Processing*, 3(4):333–341, 2009.

[14] Sungzoon Cho and Seongseob Hwang. Artificial rhythms and cues for keystroke dynamics based authentication. In David Zhang and Anil K. Jain, editors, *Advances in Biometrics*, pages 626–632, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.

[15] Nathan L Clarke and Steven M Furnell. Authenticating mobile phone users using keystroke analysis. *International journal of information security*, 6(1):1–14, 2007.

[16] Benjamin Draffin, Jiang Zhu, and Joy Zhang. Keysens: Passive user authentication through micro-behavior modeling of soft keyboard interaction. In Gérard Memmi and Ulf Blanke, editors, *Mobile Computing, Applications, and Services*, pages 184–201, Cham, 2014. Springer International Publishing.

[17] Mayank Goel, Leah Findlater, and Jacob Wobbrock. Walktype: Using accelerometer data to accomodate situational impairments in mobile touch screen text entry. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 2687–2696, New York, NY, USA, 2012. ACM.

[18] Mayank Goel, Alex Jansen, Travis Mandel, Shwetak N. Patel, and Jacob O. Wobbrock. Contexttype: Using hand posture information to improve mobile touch screen text entry. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 2795–2798, New York, NY, USA, 2013. ACM.

[19] Joshua Goodman, Gina Venolia, Keith Steury, and Chauncey Parker. Language modeling for soft keyboards. In *Proceedings of the 7th International Conference on Intelligent User Interfaces*, IUI '02, pages 194–195, New York, NY, USA, 2002. ACM.

[20] Asela Gunawardana, Tim Paek, and Christopher Meek. Usability guided key-target resizing for soft keyboards. In *Proceedings of the 15th International Conference on Intelligent User Interfaces*, IUI '10, pages 111–118, New York, NY, USA, 2010. ACM.

[21] Seong-seob Hwang, Sungzoon Cho, and Sunghoon Park. Keystroke dynamics-based authentication for mobile devices. *Computers & Security*, 28(1–2):85–93, 2009.

[22] Sevasti Karatzouni and Nathan Clarke. Keystroke analysis for thumb-based keyboards on mobile devices. In Hein Venter, Mariki Eloff, Les Labuschagne, Jan Eloff, and Rossouw von Solms, editors, *New Approaches for Security, Privacy and Trust in Complex Environments*, pages 253–263, Boston, MA, 2007. Springer US.

[23] Hassan Khan, Urs Hengartner, and Daniel Vogel. Targeted mimicry attacks on touch input based implicit authentication schemes. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*, pages 387–398. ACM, 2016.

[24] Hassan Khan, Urs Hengartner, and Daniel Vogel. Augmented reality-based mimicry attacks on behaviour-based smartphone authentication. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, pages 41–53. ACM, 2018.

[25] Emanuele Maiorana, Patrizio Campisi, Noelia González-Carballo, and Alessandro Neri. Keystroke dynamics authentication for mobile phones. In *Proceedings of the 2011 ACM Symposium on Applied Computing*, SAC '11, pages 21–26, New York, NY, USA, 2011. ACM.

[26] Sven Mayer, Huy Viet Le, and Niels Henze. Estimating the finger orientation on capacitive touchscreens using convolutional neural networks. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*, ISS '17, pages 220–229, New York, NY, USA, 2017. ACM.

[27] William Melicher, Darya Kurilova, Sean M Segreti, Pranshu Kalvani, Richard Shay, Blase Ur, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, and Michelle L Mazurek. Usability and security of text passwords on mobile devices. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 527–539. ACM, 2016.

[28] John V Monaco, Md Liakat Ali, and Charles C Tappert. Spoofing key-press latencies with a generative keystroke dynamics model. In *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–8. IEEE, 2015.

[29] Fabian Monrose, Michael K. Reiter, and Susanne Wetzel. Password hardening based on keystroke dynamics. In *Proceedings of the 6th ACM Conference on Computer and Communications Security*, CCS '99, pages 73–82, New York, NY, USA, 1999. ACM.

[30] Fabian Monrose and Aviel Rubin. Authentication via keystroke dynamics. In *Proceedings of the 4th ACM Conference on Computer and Communications Security*, CCS '97, pages 48–56, New York, NY, USA, 1997. ACM.

[31] Khandaker A Rahman, Kiran S Balagani, and Vir V Phoha. Snoop-forge-replay attacks on continuous verification with keystrokes. *IEEE Transactions on Information Forensics and Security*, 8(3):528–541, 2013.

[32] Florian Schaub, Ruben Deyhle, and Michael Weber. Password entry usability and shoulder surfing susceptibility on different smartphone platforms. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*, MUM '12, pages 13:1–13:10, New York, NY, USA, 2012. ACM.

[33] Seong seob Hwang, Hyoung joo Lee, and Sungzoon Cho. Improving authentication accuracy using artificial rhythms and cues for keystroke dynamics-based authentication. *Expert Systems with Applications*, 36(7):10649 – 10656, 2009.

[34] Abdul Serwadda and Vir V Phoha. Examining a large keystroke biometrics dataset for statistical-attack openings. *ACM Transactions on Information and System Security (TISSEC)*, 16(2):8, 2013.

[35] Deian Stefan, Xiaokui Shu, and Danfeng (Daphne) Yao. Robustness of keystroke-dynamics based biometrics against synthetic forgeries. *Computers & Security*, 31(1):109–121, February 2012.

[36] Pin Shen Teh, Andrew Beng Jin Teoh, and Shigang Yue. A Survey of Keystroke Dynamics Biometrics. *The Scientific World Journal*, 2013, 2013.

[37] Pin Shen Teh, Ning Zhang, Andrew Beng Jin Teoh, and Ke Chen. A survey on touch dynamics authentication in mobile devices. *Computers & Security*, 59(C):210–235, 2016.

[38] Chee Meng Tey, Payas Gupta, and Debin Gao. I can be you: Questioning the use of keystroke dynamics as biometrics. In *Annual Network and Distributed System Security Symposium 20th NDSS*, pages 1–6. Research Collection School Of Information Systems, 2013.

[39] Radu-Daniel Vatavu, Lisa Anthony, and Quincy Brown. Child or adult? inferring smartphone users' age group from touch measurements alone. In Julio Abascal, Simone Barbosa, Mirko Fetter, Tom Gross, Philippe Palanque, and Marco Winckler, editors, *Human-Computer Interaction – INTERACT 2015*, pages 1–9, Cham, 2015. Springer International Publishing.

[40] Emanuel von Zezschwitz, Malin Eiband, Daniel Buschek, Sascha Oberhuber, Alexander De Luca, Florian Alt, and Heinrich Hussmann. On quantifying the effective password space of grid-based unlock gestures. In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia*, MUM '16, pages 201–212, New York, NY, USA, 2016. ACM.

[41] Emanuel von Zezschwitz, Anton Koslow, Alexander De Luca, and Heinrich Hussmann. Making graphic-based authentication secure against smudge attacks. In *Proceedings of the 2013 International Conference on Intelligent User Interfaces*, IUI '13, pages 277–286, New York, NY, USA, 2013. ACM.

[42] John Williamson and Roderick Murray-Smith. Rewarding the original: Explorations in joint user-sensor motion spaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1717–1726, New York, NY, USA, 2012. ACM.

[43] Jacob O. Wobbrock. Measures of text entry performance. In *Text Entry Systems: Mobility, Accessibility, Universality*, chapter 3, pages 47 – 74. Morgan Kaufmann, 2010.

[44] Hui Xu, Yangfan Zhou, and Michael R. Lyu. Towards continuous and passive authentication via touch biometrics: An experimental study on smartphones. In *Symposium On Usable Privacy and Security (SOUPS 2014)*, pages 187–198, Menlo Park, CA, July 2014. USENIX Association.

[45] Ying Yin, Tom Yu Ouyang, Kurt Partridge, and Shumin Zhai. Making touchscreen keyboards adaptive to keys, hand postures, and individuals: A hierarchical spatial backoff model approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 2775–2784, New York, NY, USA, 2013. ACM.

[46] Saira Zahid, Muhammad Shahzad, Syed Ali Khayam, and Muddassar Farooq. Keystroke-Based User Identification on Smart Phones. In *LNCS*, volume 5758, pages 224–243, 2009.

## A Calculating Entropy of Modified Passwords

For a random password with no modifications of length n on the alphabet Σ we calculate entropy E as:

$$E_0 = log_2(|\Sigma|^n)$$

For one modification we choose a password first and then add a single modification at a random location. There are 7 possible modifications (assuming that one manifestation of each feature would be the default (e.g., pressing keys in the centre). Finally we exclude the single case where a flight time would be applied to the first character (as it does not have a preceding character to measure flight time from). This yields:

$$E_1 = log_2(|\Sigma|^n \cdot (7n-1))$$

Analogous, we calculate the entropy for two modifications by choosing a password first and then either applying two modifications on one character (15 options) or two single modifications; again excluding cases where a flight time modification would be applied to the first character.

$$E_2 = log_2(|\Sigma|^n \cdot \underbrace{((15n-6)}_{\text{2 on one}} + \underbrace{(\frac{7n \cdot 7(n-1)}{2} - 7(n-1))))}_{\text{2 single}}$$

We calculate entropy for three modifications analogously, taking into account the possibility of three modifications on one character (line 1), two modifications on one character combined with a single modification (line 2) and three single modifications (line 3):

$$\begin{aligned} E_3 = & log_2(|\Sigma|^n \cdot ((13n-9) \\ & +(15n \cdot 7(n-1) - 57(n-1)) \\ & +(\frac{7n \cdot 7(n-1) \cdot 7(n-2)}{6} - \frac{7(n-1) \cdot 7(n-2)}{2})) \end{aligned}$$

# Why people (don't) use password managers effectively

Sarah Pearman, Shikun Aerin Zhang, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor
*Carnegie Mellon University*
*spearman@cmu.edu, shikunz@cs.cmu.edu, {lbauer, lorrie, nicolasc}@cmu.edu*

## Abstract

Security experts often recommend using password-management tools that both store passwords and generate random passwords. However, research indicates that only a small fraction of users use password managers with password generators. Past studies have explored factors in the adoption of password managers using surveys and online store reviews. Here we describe a semi-structured interview study with 30 participants that allows us to provide a more comprehensive picture of the mindsets underlying adoption and effective use of password managers and password-generation features. Our participants include users who use no password-specific tools at all, those who use password managers built into browsers or operating systems, and those who use separately installed password managers. Furthermore, past field data has indicated that users of built-in, browser-based password managers more often use weak and reused passwords than users of separate password managers that have password generation available by default. Our interviews suggest that users of built-in password managers may be driven more by convenience, while users of separately installed tools appear more driven by security. We advocate tailored designs for these two mentalities and provide actionable suggestions to induce effective password manager usage.

## 1 Introduction

Despite years of searching for viable alternatives, text passwords remain as ubiquitous as they are challenging and frustrating for most internet users. Experts often recommend password managers that combine secure password storage and retrieval with random password generation. These are seen as tools that can improve account security while also improving the usability and convenience of text password authentication [40]. However, use of separately installed password managers still seems to be relatively uncommon. Previous studies have suggested that many users are not certain what password managers are, how to use them, and/or whether they are trustworthy [1, 40].

We describe a 30-participant interview study with people who do not use password managers at all, people who use password managers built into their browsers (e.g., Chrome) or operating systems (e.g., Apple Keychain), and people who employ separately installed password-manager applications. Our findings emphasize tradeoffs between convenience and security in password management and password-manager adoption, and we confirm and contextualize multiple barriers to adoption and effective usage that have been described in previous work [1]. We also highlight factors that we do not believe have been discussed previously, including confusion about the source of browser password-saving prompts and about the meaning of "remember me" options.

Furthermore, previous work has indicated that users of separately installed password managers are more likely to use unique, strong, randomly generated passwords, while users of built-in password managers may be more prone to weak passwords and password reuse. Lyastani et al. discussed that these patterns may result partially from separately installed password managers more often having integrated generators [26]. We present evidence of differences in initial motivations that may also contribute to these reuse patterns. We also provide actionable suggestions to target the three aforementioned groups of participants and induce effective password-manager usage.

## 2 Related Work

We summarize prior work on users' password habits and management choices as background for our work. We also


*USENIX Symposium on Usable Privacy and Security (SOUPS)* 2019.
August 11–13, 2019, Santa Clara, CA, USA.

describe studies that have explored adoption of password managers, as well as problems with password managers that might hinder their effective use.

## 2.1 Password Habits and Management

Studies exploring people's current password habits and burdens [15, 31, 40] provide crucial context in understanding users' password-management choices. The typical user has been estimated to have between 16 and 26 password-protected accounts in active use [11, 31, 44], and recent reports indicate that the average workplace password-manager user may have hundreds of accounts [16]. Password reuse can have serious consequences for users and organizations affected by data breaches [7, 22, 28, 30, 34, 43]. Experts thus recommend using unique, strong passwords for all accounts [21] or at least for high-value accounts [42].

However, users often struggle to remember passwords, especially infrequently used passwords [13, 40], passwords created under certain types of website requirements [36], and randomly generated passwords [45]. Users cope with these demands in part by frequently reusing passwords across multiple accounts [9, 26, 31, 44]. Along with memorability challenges, users' inaccurate perceptions of password strength and difficulty entering long or complex passwords on mobile devices also lead them to create weak passwords [27, 42]. All of these factors make a strong case for password managers.

## 2.2 Password-Manager Adoption

Security experts often recommend password managers with storage and random-generation features to help users employ strong and unique passwords without incurring memorability issues [5, 19, 20]. However, previous studies have showed password managers (particularly stand-alone applications) suffer from low adoption rates [38, 40], especially among non-experts [21, 41]. Using in-depth semi-structured interviews, we explore possible reasons for low password-manager adoption rates, as well as non-expert users' understandings and opinions regarding approaches to password management.

In 2014, Stobert and Biddle conducted 27 semi-structured interviews to examine the "life cycle" of password use. They found the rationing of effort to be a central theme in users' password-management choices. Almost all of the participants in this study reported using password managers built into web browsers, but none were using separately installed password managers [39]. Our study uses a similar interview approach to explore themes including users' strategies for creating and managing their passwords, with the intuition that the password ecosystem may have become more complex for some users since 2014. Furthermore, we intentionally sought out users and non-users of password managers in order to examine what factors drive them to adopt or not adopt these tools.

Alkaldi and Renaud conducted a web survey as well as analyzed reviews for password-manager apps in the Google Play Store and broadly listed many observed reasons for adoption and non-adoption [1]. Fagan et al. surveyed 248 people on MTurk to probe their reasons for using or not using password managers as well as their emotions associated with the usage of password managers. Similarly to the work by Alkaldi and Renaud, this survey asked participants, in an open-ended question, why they chose [not] to use a password manager. This work also provides a number of broad reasons including "security concerns," "lack of need," and "lack of motivation/time." They found that password-manager users tend to regard "convenience" and "usefulness" as their main reasons for adoption, and "non-users" are more likely to feel suspicious compared to "users" [10]. Similarly, Aurigemma et al. conducted a survey with 283 undergraduate students who reported they did not adopt password managers because they lacked time for installation, the sense of urgency, or the awareness of how password managers worked [3]. Alkaldi and Renaud also conducted another study to test an Android application to recommend password managers to users and found that such an intervention may be most effective when it appeals to users' autonomy (sense of control) and relatedness (sense of community with others) [2]. Our study complements these studies with interviews to present a more comprehensive picture of why and how people arrive at their decisions of using or not using password managers.

## 2.3 Problems with Password Managers

In 2006, Chiasson et al. studied two password managers and identified several usability issues caused by: i) users' incorrect or incomplete mental models of the tool, and ii) users' feelings that they did not need password managers and unwillingness to hand over control [6]. In 2011, Karole et al. evaluated the usability of password managers running on a website, on a mobile device, or on a USB device. They found that non-technical users preferred to manage passwords on their mobile devices rather than relinquish control to a web-based password manager [23].

Besides the aforementioned usability issues, prior work has highlighted some security vulnerabilities in password managers, although experts generally still consider password managers a net positive [12, 18]. Li et al. identified various vulnerabilities associated with five popular web-based password managers [25]. Silver et al. revealed risks of the auto-fill functionality provided by many popular password managers [37]. Research has also indicated problems with local data security. Gray et al. revealed that unencrypted password data could be found in temporary folders [17]. Belekno et al. [4] and Gasti et al. [14] described risks that exist when attackers possess physical access to users' devices or password databases.

In 2018, Lyastani et al. collected *in-situ* password data of MTurkers through a Chrome plug-in. They found that

Chrome's autofill seemed to encourage password reuse, while users of LastPass' integrated generator tended to have stronger and less-reused passwords [26]. Our study provides more insight about the mindsets of these types of users, as well as those of users who do not use any password tools.

## 3 Methodology

We conducted 30 semi-structured interviews to probe participants' current password behaviors as well as their attitudes, beliefs, and understandings surrounding password creation and composition, account security, and password management and storage.

### 3.1 Recruitment

We recruited participants from Pittsburgh, PA using both online outreach (posts on Craigslist, Reddit, and Facebook) and offline strategies such as posting flyers on community bulletin boards. We used purposive sampling to ensure that we interviewed participants who used a variety of password-management strategies, including non-technological approaches (e.g., writing passwords in a notebook), computer-based approaches that did not involve password-specific software (e.g., saving passwords in an Excel spreadsheet), password managers built into web browsers, password managers built into operating systems (e.g., Apple Keychain), and separately installed password managers. We stopped recruiting when our sample included multiple participants from each of the above categories. We also sought diversity in age, occupation, and level of technical knowledge.

Potential participants were asked to take a short screening survey to confirm eligibility (age 18 or older, able to speak English), availability, estimated number of internet accounts, types of devices used (laptop, desktop, smartphone, tablet, other), primary operating system(s) for those devices, password-management strategies, past experiences with compromised accounts or data breaches, and basic demographics.

### 3.2 Interviews

This research was approved by our university's institutional review board. Participants completed a consent form, were given the opportunity to ask the researcher questions before beginning, and were instructed that they could stop the interview at any time or decline to answer any question. The interview script is shown in Appendix B.

Participants were interviewed in person on our institution's campus. One primary researcher was present for all 30 interviews. A second researcher assisted in some interviews. Each interview lasted approximately one hour. At the end of the interview, each participant filled out a brief demographic survey. Participants received a $30 Amazon.com gift card.

With participants' consent, all interviews were recorded, and the recordings were then transcribed by a commercial transcription service. Participants were asked not to share actual passwords or other identifying information, but when participants did mention details that seemed likely to be sensitive or identifying, the recordings were trimmed of those details before being sent to the transcription service.

### 3.3 Analysis

Interview transcripts were analyzed using inductive coding. An initial codebook was created based on the interview script and early interviews, and two researchers collaborated iteratively to improve the codebook throughout the coding process.

Five of the 30 interviews were coded by both researchers to ensure inter-rater reliability. The average Cohen's kappa, a commonly-used statistic reflecting agreement among coders, was 0.84, which denotes a very high level of agreement [24]. All coding discrepancies in these five interviews were discussed and reconciled. The remaining interviews were coded independently (10 by one researcher and 15 by the other); however, the researchers met regularly throughout the process to discuss any perceived ambiguities in the coding of particular data points as well as any necessary changes or additions to the codebook. These methods were deemed sufficient given that the results reported are qualitative and exploratory.

The final codebook contained 309 total codes across 59 categories. Most categories reflected a particular question or topic from the interview script and the types of responses observed to that question: for example, one category of codes called "current password management" included codes such as "physical notebook" and "browser password manager." A "miscellaneous" category also captured certain high-level themes that were observed repeatedly, e.g., "device sharing."

### 3.4 Demographics

We interviewed 19 users who identified as female and 11 who identified as male. Four were 18–24 years of age, nine were 25–34, eight were 35–44, five were 45–54, three were 55–64, and one was between 65 and 75. Most users were highly educated: 21 had bachelor's degrees, and nine of those also had graduate degrees.

Nine users worked in technical fields. We sought to interview a more representative sample, but we encountered difficulty in recruiting users of separately installed password managers when we excluded those with technical backgrounds. Only two participants self-identified as security professionals.

### 3.5 Limitations

We emphasize that this study is qualitative and based on a purposive sample, and we are not making any quantitative comparisons or claims. Our population sample skews female and

young, and most of our participants had bachelor's degrees or higher. We also had a disproportionately high percentage of participants with technical backgrounds, largely because we found it difficult to recruit users of separately installed password managers who did not have technical backgrounds. We do not claim any generalizable statistical findings from this study: our goal is to describe some of the user types to consider when designing and marketing password managers, as well as some of the barriers to adoption and effective use.

Due to our screening survey and purposive sampling methods, participants likely came to the interview believing that we were security researchers interested in password-management tools. This could raise concerns about priming and the Hawthorne effect, i.e., that participants might indicate more affinity for password managers than they actually had. Nonetheless, many participants still told us about habits that they knew were not considered secure and about reasons that they did not like or did not want to use password managers.

We also acknowledge that users may not have been able or willing to self-report their password habits accurately in all cases. However, in complement to existing in-situ data, these self-reports offer crucial insights regarding the mindsets underlying users' observed behavior.

# 4 Results

Many of the users interviewed had complex password strategies, including multiple password-storage methods. However, we have categorized interviewees based on whether their primary approaches to remembering passwords depend on non-password-specific methods, built-in password managers, or separately installed password managers. Here we describe how these groups characterized their current password habits and their attitudes towards password-management options.

## 4.1 Password-Management Approaches

**Approaches Not Involving Password-Specific Tools** The first group of participants that we will discuss used approaches that did not involve any type of tool designed specifically for password management as their primary method. This includes memorizing passwords, writing passwords (or hints to passwords) on paper, sending oneself emails or voicemails containing passwords, listing passwords in unencrypted computer files (e.g., a Microsoft Word file), or listing passwords in note-taking applications (e.g., iPhone Notes app). Some participants also relied heavily on the ability to reset forgotten passwords. Nine interviewees were in this group.

Some of these participants had used password managers incidentally or in the past. P27, for example, reported that he was able to log into a few apps on his Android smartphone with his fingerprint, suggesting he was likely using Google Smart Lock to a limited degree. However, his primary strategy was to memorize his passwords. P28 also had some passwords that were saved in the browser on an infrequently-used home computer, but she reported that these were outdated and that she did not save passwords when prompted anymore.

**Built-In Password Managers** The second group of participants discussed below primarily used password managers built into browsers (e.g., Apple Safari, Google Chrome, and Mozilla Firefox) or operating systems (e.g., macOS Keychain Access & iCloud Keychain, or Google Smart Lock for Android and ChromeOS) to store and autofill some or all of their passwords. The distinguishing feature of these tools versus other password-management tools discussed below is that they are present in the browser or the operating system as standard features. To access these tools, users may need to install browsers that are not built into their operating systems, but they do not need to install additional password-specific applications or extensions. Twelve belong in this group.

In many cases, browser-based and operating-system-based tools from the same company are integrated with each other: for example, passwords saved in Safari may be viewed in Keychain Access on macOS and may be set up to sync to the cloud and to iOS devices using iCloud Keychain (all Apple products). For this reason, we discuss all of these built-in tools together rather than distinguishing browser-based tools from operating-system-based tools.

**Separately Installed Password Managers** The third group of participants that we will discuss are those who used some type of separately installed password manager, i.e., tools that are not built into browsers or operating systems and must be installed as separate applications and/or browser extensions. We interviewed seven users in this group: four users of 1Password, two users of LastPass, and one user of KeePass.

**Other Approaches** Two participants were difficult to place in the aforementioned categories. P29, who described his approach as "security by obscurity," reported using a combination of memory, mnemonics, and browser password storage to handle passwords on a routine basis, but he mainly stored his password list using an application called Cardfile that he described as a "Windows 3.1 executable." He updated this file manually on a home machine running Windows XP. P21 created his own encrypted file using PGP to store his passwords, and accessed them through SSH when at work.

## 4.2 Current Password Habits

**Account Numbers** When asked how many password-protected accounts they had, almost all participants (except for five users of separately-installed password managers) gave answers under 100. Most of the participants who do not use password managers gave an answer under 50, and two reported having more than 50 but under 100 accounts. For users

of built-in password managers, estimates most commonly ranged between 15 and 50 accounts. For users of separately installed password managers, the five with technical jobs all reported having well over 100 password-protected accounts, with one reporting having over 1000 accounts. The remaining two users reported having 20-50 accounts.

**Password Reuse** Of participants who do not use password-management tools, seven indicated multiple risky password habits, including heavy reuse of passwords and few or no unique passwords. However, one user reported that none of her passwords were reused exactly but that she did reuse substrings when creating passwords (although always in different positions in the passwords). Another user reported that most of his passwords were unique but also reported that his strategy for creating passwords was to use words related to "kids," "names," cities, or states, and then add numbers, which suggests that his passwords may have been highly guessable.

Only one participant who primarily relied on built-in password managers specifically reported efforts to have unique passwords for all important accounts. About half of the other built-in password-manager users indicated heavy reuse of one password for all of their accounts. The rest employed various strategies to decrease the extent of their password reuse: some applied a tier system, using a unique password for accounts of similar importance, and some tried to have unique passwords for important accounts but still engaged in insecure practices like reusing parts of their passwords or using memorable personal information in their passwords.

Of the seven participants who use separately installed password managers, all but one reported switching to password manager use gradually, with only one participant (P23) reporting an effort to change all passwords to randomly generated, unique passwords at the start of using a password manager. P23 reported that this took at least five hours over the course of about three days to migrate the 40 accounts that he had at the time, which was three to four years prior to the interview. (At the time of the interview, he estimated that he had about 300 accounts.) Another participant (P19) did not commit to changing all of his passwords at the beginning. After one of his reused passwords was phished, he updated hundreds of his passwords to be unique and randomly generated.

**Password Generators** Only one of the users who relied primarily on built-in password managers described using password-generation features. P10 reported using Safari's password-generation tool to create random passwords for important accounts. She used Apple's Keychain functionality to record and fill these passwords. (She described reusing a weaker password across some low-value accounts.)

She indicated a recent account breach as the impetus for this strategy: previously, she had been reusing many of her passwords, and then an attacker gained access to a department store account as well as her email. She described this as a traumatic experience that caused an immediate desire to change her habits:

> All of my information was just taken. It was awful, so I'm having to get a new debit card for everything, having to get new credit cards... That's when I realized that I needed to reevaluate. That's when I changed every single password that I had to random digits. I didn't even think twice. I was like, "Something needs to change, and it has to change on my end."

At the time of these interviews, Safari did offer a password generator, but the six other participants who used Safari on some of their devices did not report use of this feature. Google Chrome began to roll out Chrome 69, which included a new password-generation feature, in fall 2018 [8, 33]. Some of our interviews were conducted after the release of Chrome 69, but no Chrome users mentioned awareness or use of the feature.

All seven users of separately-installed password managers reported using randomly generated passwords when creating new accounts, and most of these participants used unique passwords for newly-created accounts (with the exception of P22, whose strategy is described in more detail below). P30 reported using websites to generate random passwords before realizing that LastPass had that ability.

P22's strategy was distinct from that of the other participants. First, he did not use the password generator built into 1Password: he instead preferred to use other generators such as one offered by Symantec, which he reported was simply a matter of "habit." Second, he did not use the generated passwords in their original form, but instead made changes of his own by adding characters in the middle and/or removing some characters before saving the passwords in order to make them "more secure" and "more random."

Additionally, P22 reported that he reused passwords in tiers rather than storing unique passwords for accounts. For example, he reported that he might use the same password for all social media accounts. He did this out of worry that he might not have access to his password manager in certain situations or if borrowing someone else's device.

**Master Passwords** The seven participants who use separately installed password managers employed a number of different approaches to deal with their master passwords. All but one of those seven (P20) reported using a unique password as their master password. P20 reported that her master password was one of her three heavily reused passwords. Some participants (P18, P23) reported using passphrases as their master passwords, like "a quote from a movie" (P23) or "a sentence that doesn't make sense" (P18). Some (P17, P19, P22, P30) indicated that their master password was randomly generated. P17, P19, and P22 used 1Password, which prompts users to memorize a randomly generated master password when creating an account. P30, who used LastPass, reported

using a website to generate a random master password, but was unable to memorize it. She kept written copies at home, at work and saved it in a draft inside her email account. P20 and P30, who engaged in unsafe practices regarding their master password, did not have technology-related degrees or technology-related jobs.

## 4.3 Experiences of Participants Not Using Password Managers

Nine participants were not using any password-specific technology or tools to help them manage their passwords. Some participants without password managers were satisfied with their password-management approaches, but others were concerned about password security or found their current approaches inconvenient.

**Satisfaction with Current Method** Some participants who were not using password-specific tools liked specific aspects of their current password-management methods, which may inform efforts to target password-management tools to people who currently do not use such tools.

P11 and P27 noted that password reuse made it easy for them to remember their passwords. P11 noted that this was due to always using a default password. P27 noted, "it's easy because I've been using the same variations for a while... It's like my phone number. I know it without thinking about it."

P4 liked keeping a copy of her passwords outside of her browser because she felt that passwords stored in a browser password manager might be lost—e.g., if the IT department at her workplace had to wipe a computer during troubleshooting.

Some participants liked having control over how their passwords were organized. P4 kept them in alphabetical order with notes about the account they belonged to, while P7 kept them grouped by type of account.

P12 liked keeping his passwords in a list in a note on his phone because he could bring this list with him anywhere. He mentioned not being aware of how else he could have access to his passwords on the go.

**Dissatisfaction with Current Method** Some participants in this group did like aspects of their current method of storing passwords. However, five participants were dissatisfied with their current password-management methods, and several participants described negative aspects of their current methods, including recall difficulty, disorganization, access problems, and potential security risks.

Some participants emphasized that it was difficult to recall their passwords. P28 talked about having to reset passwords "constantly" due to forgetting them. As mentioned above, P27 said that it was generally easy to remember passwords that he had reused for a long time, but he also encountered difficulties when trying to remember what password variation he had used for a particular websites' password requirements:

> It can be hard because I don't remember if a website wanted me to have a capital letter or if they wanted me to have a symbol or if they wanted at least 12 characters.

While some participants liked how they organized their password lists in paper notes or in files, P26 felt like her system of using her memory as well as writing a few passwords down was "disorganized":

> I put a lot of energy into trying to remember what's what. I'm like, "I could be doing something else with that energy."

P4 described problems with accessing accounts when away from where her password list file was stored. She kept a Microsoft Word file on her computer and also kept a printed copy, but she did not carry a copy with her on paper or on her phone. She felt that carrying passwords with her was risky.

> I could [keep the list on my phone]. And I could email it to myself too but... I feel like you're putting more risk when you do all those things. I mean, what's the point of having passwords if you're gonna carry them on your body and say, "Hey, this is my password." You know. But yeah, I can't always access them, to be honest.

Some participants who stored passwords in files or digital notes expressed concern that an unauthorized user of one of their devices might be able to access these password lists. P4, who kept passwords in a Microsoft Word file, said, "It's on the computer, which I know is really bad. But I don't name that 'passwords' on the computer. Just in case somebody got on my computer."

P12, who stored passwords in a note that was saved only on his iPhone, was concerned about what this would mean if his phone was stolen or used by someone without permission: "I know it's dumb, but I save them in my phone in my notes, so if someone has my phone, I'm through, right?"

## 4.4 Barriers to Adoption Among Participants Not Using Password Managers

Some users who were not using password-management tools were simply unaware that they existed. Additional barriers to adoption expressed by this group included security concerns, believing they did not have much to protect, concerns about the single point of failure, or past negative experiences with password managers.

**Awareness** Some users in this group did not believe they were using the best password-management methods, but they also were not sure if better options existed. P4, for example, said there could be a better way, but she was not aware of one, suggesting that sheer awareness of password-management

tools is the primary adoption barrier for some users. Similarly, P14 wished for "an easier way to remember passwords, like a universal-type system."

**Security** Six out of the nine participants expressed various concerns about the security of password tools.

Some expressed concern or lack of knowledge about the security of password managers. P12 noted that he would need to learn more about their security. P11 wondered if password managers were "really safe and secure" and described generally preferring "pen and paper" due to being "leery of technology." P11 wanted to know where and how password managers stored passwords, as did P27.

Some participants had considered using browser password features but were uncertain about their security. P26 described declining browser prompts to save passwords because "it feels insecure." P28 had stored passwords in Chrome in the past and said it felt easy, but she stopped because she was not sure whether Chrome stored passwords securely.

P11 also described confusion about who or what was prompting her to store passwords in her browser:

> I don't know if it's Google that's asking me, I don't know if it's the website that's asking me... that would be one reason [to not save passwords].

Some participants were reluctant to use password managers due to concerns about specific types of attackers, including external attackers (i.e., "hackers"), employees at password-manager companies, and other users (authorized or unauthorized) of their devices. P14 worried about important accounts being hacked if their passwords were saved:

> I'm afraid that if I use it, I might be sorry in the end. I might have a hacker get into my system. You know? Some people have nothing to do and they'll just hack into people's computers for no reason. This is just kind of like insecurity.... I use it [Chrome's password manager] for certain accounts... but I don't think I'd use it for my email or my bank...

P27 expressed concerns that employees at a company offering a password manager might be able to decrypt and access his passwords. P27 was also concerned about other users of his device accessing his accounts, as were P4 and P11.

**Not Enough to Protect** Three participants felt that they did not have enough accounts or that their accounts were not valuable enough to require a secure password-management tool. P5 and P27 felt that they were able to remember their passwords without help.

> My life on the Internet is not that complicated with my 15 passwords that I can more or less remember and my little book. But if it were to get anymore complicated than that if I were to have dozens of accounts, then yes, I would.... I would think that there's a far superior way to deal with passwords if you were using a huge number of them. (P5)

P5 and P12 felt that their accounts were not sufficiently high-value to require extra security like that offered by a password manager. P5 said that she might use one if some accounts were protecting more important financial information. P12, who mentioned he used prepaid credit cards for online purchases instead of cards connected to bank accounts, said he might consider using a password manager if he had bank accounts or medical records to protect.

Some participants had trouble conceptualizing the mechanisms of "hacking" or the statistical risk of their own accounts being compromised. P7, for example, when discussing his use of a website from which 340 million records were exposed, believed that the probability of his own information being exposed was "one in 340 million" and thus was not very concerned: "I don't care... I'll take that risk anytime. I'm more likely to walk out of here and get run over by a car, right?"

**Single Point of Failure** Some participants worried about storing all of their passwords in one place. P27 was concerned about security: "If somebody found out the way that they encrypt it, they would be able to get access to all my passwords at once instead of one of my passwords." He also worried that it would be difficult to create new passwords if all of his passwords needed to be changed. P11 and P26 were afraid of losing access to all of their passwords due to a forgotten master password or other problem.

**Past Negative Experiences** Three participants in this group had negative experiences with password managers before. They were unable to reliably store their passwords using those tools. P4 reported that Google Chrome had sometimes saved her passwords incorrectly, such as with a lowercase letter rather than uppercase. She also recalled losing passwords stored in her phone after clearing the browser cache. Similarly, P14 said that she had tried to use Chrome's password manager in the past but that "it mostly doesn't work out." She said it saved her personal information, such as her address, but that it did not save her passwords: "Even though it says it's going to save it, it doesn't.... I wish it would save more."

P28 had experienced trouble with password managers in the past because she changed her passwords frequently: she described cases in which the browser would fill in the old password and cause failed logins. P28 had also tried briefly to adopt LastPass, but she found master password creation to be a "hurdle": "the last thing I would want is to have a database with all your passwords with a dumb [master] password."

## 4.5 Experiences of Users of Built-In Password Managers

Of 12 participants who were using a password manager built into a browser or operating system, half (six) reported that they were not satisfied with their current password-management choices. P3 said, "I know there should be a better way." Two participants were uncertain and expressed wishes for an easier or safer way to manage passwords.

Participants generally reported adopting built-in password managers due to seeing prompts or for reasons of convenience, and the aspects of these tools that they liked included convenience-focused features such as autofill. These participants did not report choosing these tools for reasons of security. In many cases, they believed that their password habits were risky—probably correctly, since most of them reported significant password reuse—but did not feel motivated to change those habits and were not aware of features such as password generators that would assist them in doing so.

**Likes** Almost all participants emphasized liking autofill. P16 reported that it saved time, and P24 enjoyed the convenience of not having "to always type in a password."

P25, who was more familiar with Chrome's features than other users, liked that Chrome allowed her to sync passwords across devices and protect her passwords with multi-factor authentication.

In addition to any comments about password-management tools, half of the participants in this group mentioned that they liked reusing passwords because it allowed them to remember their passwords easily. P3, for example, said:

> It's a no-brainer. I don't have to think about it, I just automatically do it. I'm 68 years old and I don't want to have to remember more than I have to.

**Dislikes** Four participants (P3, P6, P15, and P24) expressed concerns over other people having access to passwords that they saved using built-in tools. P6, for example, was worried that her child's friends, who she said borrowed her computer sometimes while visiting her house, might make online purchases using her accounts.

One other complaint is the accessibility of passwords on devices that the passwords were not saved to (P2, P6).

> Auto-fill is awesome. I'm starting to rely on that more and more. It's just if you don't have that device... and you're somewhere else when you need it, that's the only downfall I can see. (P2)

Ten out of the 12 participants did not know or believe that their password manager allowed them to view a list of all passwords, although Chrome, Firefox, and Safari do offer this. Four participants (P2, P6, P9, P13) emphasized that they would like the ability to view all of their saved passwords.

P15 pointed out that sometimes the built-in tools did not update her passwords when she changed them.

**Factors in Adopting Built-In Password Managers** No participants mentioned security or unique passwords or randomly generated passwords as a reason for adopting a built-in password manager. Users in this group all focused on prompts, convenience, or memory limitations as adoption factors.

Nine out of 12 participants in the group remembered receiving prompts from the tool offering to save passwords for them. Seven quickly accepted these prompts, and these users gave no other particular reason for their adoption of the tool. Two participants (P1 and P25) were slower to accept those prompts. P1 mentioned that she finally decided one day to click "yes" when she was feeling "lazy." P25 said that she was skeptical about letting Chrome save her passwords but that she felt more comfortable after hearing that her friends liked Chrome's password manager and found it to be convenient.

Six participants emphasized convenience as a reason for using the built-in tool, emphasizing benefits such as "faster" log-ins. P2 mentioned that forced password changes made it hard for him to remember all of his passwords on his own.

## 4.6 Barriers to Effective Use Among Users of Built-In Password Managers

Users of built-in password managers often adopted them for reasons of convenience or due to seeing prompts. Accordingly, they often did not use them in effective or secure ways because they (perhaps incorrectly) believed themselves to be at low risk or because they did not have sufficient knowledge.

**Risk Assessment** Like many of the participants who were not using password-management tools, the 11 participants who stored reused passwords in their built-in password managers were often reluctant to change their habits because of a lack of personal experience with account compromise, a perception that they were at low risk of account compromise, or a belief that an account compromise would not have important negative effects.

Eight of the 11 participants who stored reused passwords in their built-in password managers acknowledged that their password habits put them at some risk. However, P1, P8, and P25 said they were not likely to change their behavior since they had not experienced negative consequences so far.

> Yes, it's a bad idea to have all the passwords like to be very similar, but I think that because I haven't been personally affected by someone... hacking me or changing or grabbing my info... I'm less inclined to change the way I manage my passwords. (P1)

P10, the only built-in password manager user who was using randomly generated passwords, reported that she also

had not believed herself to be at risk of account compromise until she experienced it first-hand:

> Not really, because I hadn't really experienced anything like [being hacked]. It was just...a complete eye-opener. I was like, "Nobody's ever going to hack anything, nothing." Completely naïve.

Another two participants (P3 and P9) did acknowledge risk in their behavior but also felt that their accounts were not of much value. Both of them were aware of password managers, and one had even used a separately installed password manager previously. P3 explained that she only shopped online for "tiny things" and that her wife, who handled most of their finances, had more reason to be concerned about passwords.

P9 claimed that she did not have much to lose: "[T]hey can have my bank account; there's not that much money." P9 was one of a few participants who mentioned that they were more careful with other people's information than their own: she described having better password practices when she was working in a university job where she was responsible for research data on her computer.

**Lack of Awareness or Knowledge** As we find in participants who do not use password managers, lack of knowledge remains an obstacle for built-in password manager users. Four out of 12 participants were not aware of the term at all, and only one (P16) expressed awareness of separately installed password managers.

Furthermore, as discussed in Section 4.2, only one built-in password manager user used or was even aware of the password generation feature included in their password manager.

A lack of information also caused some of these participants to be reluctant to consider separately installed tools. After the interviewer offered a description of available password management options, three participants (P2, P3, and P13) expressed concerns about password managers from unfamiliar companies. P2 said that he might use a tool offered by a known company like Google but that he would be "leery" of a "new-name company":

> Any kind of third party scares me a little bit. I don't know who or what is doing that part of it, and what information is shared out there.

## 4.7 Experiences of Users of Separately Installed Password Managers

Five separately installed password-manager users (P17, P18, P19, P22, and P23) reported that they were satisfied with their password managers, but the two users without technical backgrounds (P20 and P30) had less positive perceptions. P20 referred to her password manager as "a pain in the ass" but was not convinced a better solution would ever exist.

> There's not going to be a better solution, I'm going to not store them occasionally, and I'm going to have to call and get someone to reset them... That's just the price that we pay to keep those things.... I'm in public health, so we think about prevention instead of treatment, and this is the kind of thing where you're preventing things that are happening, but you're not seeing any rewards for it. So if I don't get hacked, I don't have a party because I don't get hacked. Whereas if I get hacked... I'm assuming that's a really negative experience.

**Likes** P18 and P23 appreciated no longer needing to memorize passwords. P17 and P30 also referred to generally liking that their passwords were stored or saved. A few participants also mentioned that they used their password managers to store information other than passwords. P22 found the ability to store SSH keys particularly useful.

P19, P23, and P30 liked being able to generate random passwords. P19 specifically liked having passwords that were not vulnerable to dictionary attacks and that would likely be slow for an attacker to crack. P30 also specifically liked that the password manager helped her use unique passwords.

P18 (a KeePass user) and P22 (a 1Password user) specifically liked having a desktop client. P23 liked the ability to sync across devices. P20 appreciated the portability of this system, or the ability to have passwords available "on the go."

P19 mentioned the ability to fill passwords without typing as an important feature, but he also specifically liked 1Password's implementation: "It doesn't pick random domains to fill in for a phishing website or something like that." [1]

**Dislikes** P17 said, "From a user experience standpoint, [1Password] is a mess," noting that it often did not save usernames or passwords correctly. She also reported that 1Password would sometimes fill a password without filling the corresponding email or username and would try to submit these incomplete credentials.

P30 said, "the [LastPass] app sucks," also citing issues with the browser extension not logging into websites correctly. P20 also encountered conflicts between Google Chrome's password manager and the LastPass browser extension: since she had saved some passwords in Chrome in the past, Chrome would still prompt her about filling or saving passwords.

P23 cited the difficulty of entering long, randomly generated passwords into certain devices that were not compatible with the password manager, such as gaming consoles or Roku streaming devices. P17 also mentioned that she would prefer to have shorter, easier-to-type passwords for certain

[1] 1Password's password filling functionality is different from most password managers in that it requires user instruction to trigger the filling of a password. 1Password prefers not to refer to this as "autofill" and has made this choice for security reasons [32].

frequently-entered passwords, entry on mobile devices, or cases where the password manager was unavailable:

> I recently had to factory reset my phone, and I had to log into my Google account. And, if I hadn't had that password in my brain, then I would have to like go to my laptop, then open up 1Password, and then read this like 20 character string, and then type it... it's always so annoying.

P23 mentioned that 1Password's generator did not offer enough control to fit some websites' password requirements.

P18, a KeePass user, reported that it was "annoying" to have to sync his password database manually. He also mentioned KeePass's lack of cloud storage, and he noted that it was not easy to access his passwords from his Android phone. (At the time of writing, KeePass is available natively as a Windows client or portable application. To run KeePass on other operating systems such as macOS, Linux, or Android requires using an unofficial ported version [35].)

P20 and P30 feared forgetting their master passwords. P20 said, "I don't know if you can reset it if you forget it.... and there's something scary about that."

**Adoption Motivations** P20 and P30 cited memory limitations as a primary motivator. P30 had encountered memory difficulties after attempting to use unique passwords.

P18, P19, P20, P23, and P30 described a broad desire for increased security as an important motivator. P30's password-security concerns were heightened due to volunteer work in which she was responsible for other people's data. P22 also cited a specific desire to avoid typing in passwords manually (implying concerns about keyloggers) and a belief in password managers' use of "modern cryptography in encryption."

**Information Sources** P17, P18, and P19 gained awareness of password managers from working in IT or technology. P17's company encouraged password-manager use and paid for employees' subscriptions to premium versions of popular password managers, and P17 chose 1Password because IT staff at her company recommended it.

P20 specifically remembered hearing a story about password managers on NPR. P23 listed a number of possible places where he might have first heard about password managers, including Reddit and Hacker News.

# 5 Discussion

Our findings emphasize tradeoffs between convenience and security in password management and in password-manager adoption. We confirm and contextualize barriers to adoption and effective usage of password managers that have been covered in past work, while introducing additional factors. We also provide actionable suggestions to induce effective password-manager usage targeting three groups of users.

## 5.1 Security vs. Convenience

**Tradeoffs and Effort Rationing** A consistent theme that emerged from password-manager users and non-users alike is the tradeoff between security and convenience. Many of our participants reported making compromises about security in order to ration their efforts—even participants who were relatively concerned about security and who had fairly secure practices overall. Our findings echoed Stobert's work on password life cycles, which reported that "effort rationing" was a primary motivator of various password-related habits [40]. Multiple participants mentioned following recommended secure practices for higher-stakes accounts (e.g., financial), while employing reused and/or weak passwords for lower-value accounts. Many also mentioned reusing passwords in tiers to ration efforts.

Furthermore, our study extended this line of reasoning to users of separately installed password managers, who were not present in Stobert et al.'s study. Password managers were sometimes described as solutions that saved memory effort and/or time, but in other cases, they were described by our participants as tools that required additional effort. When these tools do not function as intended, we see users ration their efforts by circumventing these inconveniences and resorting to other methods, which are often less secure. Users of separately installed tools would choose not to randomly generate passwords but reuse old weak passwords when user-interface problems made saving logins difficult, or would email passwords to themselves when syncing was not available. Users of built-in tools would resort to recording passwords on paper or in text files if they could not trust their browsers to save passwords reliably. We believe that there is a need for better user-experience design and thorough usability testing, especially long-term user studies for corner cases. We also observed that users without technical backgrounds may encounter more problems in their use of separately installed password managers, calling for tailored design and usability testing targeting non-expert users.

**Motivations** Lyastani et al. observed lower password strength and more reuse among users of browser password managers than among users of separately installed tools. The authors suggested that browser password managers, by not integrating generators with normal password creation workflows, potentially exacerbated password reuse [26]. Our results suggest that users of built-in password managers and users of separately installed password managers often have fundamentally different motivations and that differences between password-manager interfaces are not the sole cause of the reuse patterns Lyastani et al. observed. The majority of users of built-in password managers began saving passwords due to prompts or convenience, while users of separately installed tools emphasized security concerns, limitations in remembering unique passwords, and features such as generators

and strong encryption.

However, findings from our interviews do confirm Lyastani et al.'s suggestion that having a generator integrated into the workflow for password creation is beneficial. Many users of separately installed password managers found the generator convenient, while the majority of users of built-in password managers were not aware of the feature. It seems constructive to continue adding password generation features to existing browser password tools, as Apple (and, more recently, Google) have done. Furthermore, given how many users of built-in password managers adopted those tools due to prompts, we suspect that similar prompts towards password generation might be effective. We suggest that future work investigate what types of nudges are more likely to lead to password generator adoption (or other ways to improve password habits) among these more convenience-focused users.

## 5.2 Factors Driving Adoption

Our study provided additional evidence supporting some of Alkaldi and Renaud's findings about the adoption of separately installed password managers [1]. In particular, we found evidence of the importance of: subjective norms and social influence (supported by our findings of workplace influences on adoption of these tools), time-saving and memory benefits, past experience with security breaches, and perceptions of increased security.

Some of our findings, however, added complexity to Alkaldi and Renaud's results around adoption factors. Some of our users of separately installed tools, like theirs, reported that syncing was useful and desirable, but others found it not secure and/or not useful, implying that password-manager makers might want to continue to offer both cloud-based options with syncing and locally stored options that do not sync automatically. Furthermore, given their different backgrounds and needs, some participants found password managers effort-saving and easy to use, while others found them frustrating and unreliable, emphasizing that password managers must be designed to be usable for non-experts.

Our results also lend nuance to Fagan et al.'s findings that convenience and usefulness, not security, were the primary reasons for password-manager use [10]. Fagan et al. did not distinguish between users of built-in and separately installed password managers. Our results suggest that convenience and usefulness are indeed paramount for many users of built-in password managers, but that security is often a primary motivator for users of separately installed password managers.

## 5.3 Barriers to Adoption and Effective Use

Our work confirms and contextualizes many factors discussed by Alkaldi and Renaud as leading to rejection of separately-installed password managers [1], including lack of awareness, not enough passwords or important data, and concerns about security. Participants not using password managers and participants using built-in tools reported common themes like risk assessment and lack of awareness or knowledge.

We found that certain themes from Alkaldi and Renaud's work were especially salient in the discussion of password generators in our findings. In particular, participants who discussed not wanting to use randomly generated passwords often did so because they believed being able to "master" and memorize their passwords was important. P26 indicated that not knowing her passwords would feel like giving up control. P25 emphasized the importance of being able to access her information at all times and asked, "What's the point of creating a password if you can't remember it yourself?"

We also found some specific barriers that we do not believe have been discussed in other work. First, we found gaps in some participants' underlying understanding of websites, browsers, and password saving that precluded making informed decisions about using password managers. Some participants who had received prompts about browser password-saving features were unsure where the prompts were coming from: the browser, the website, or the computer. Participants were also sometimes uncertain where passwords would be stored or whether employees of the company making the password manager could see their passwords. Password managers should help people understand where their passwords are stored and what security measures are taken to protect their passwords in order to help them make informed choices.

Some participants also expressed confusion about password managers because they recalled checking "Remember me" options on login pages and not having their passwords saved. Multiple participants were confused about whether this option (which might normally either create a persistent login with a cookie or cause a username to be prefilled for subsequent page visits) was part of the browser password manager. These confusions lead users to lose trust in the reliability of those browser password managers, which propels them to resort to other insecure methods as mentioned in Section 5.1.

A few participants echoed concerns about having "all eggs in one basket," i.e., a single point of failure, which also appeared in Alkaldi and Renaud's survey responses. Some users of separately installed password managers also acknowledged this risk, but they felt that security benefits such as strong and unique passwords outweighed those concerns. Clear, accessible information about how password managers store and protect passwords may offer non-expert users a more accurate understanding of password managers' risks and benefits. Password managers that offer multi-factor authentication might also increase confidence for some of these users.

**Targeting Non-Users** Some people who were not using any tools to manage passwords were simply unaware of the existence of such tools. Some were uncertain about the security of those tools to protect from external attackers, thieves or others with unauthorized access to their devices, or other

(authorized) users of their devices. We suggest that further research might explore how advertising, education, or browser prompts could target those who are not currently using password tools, who may often be individuals who have less experience or expertise with technology. Accessible information should be offered that emphasizes not only security against remote attacks but also features that allow the user to control whether passwords are accessible to others with physical access to the device.

Furthermore, some participants using their own notes emphasized that they liked being able to sort passwords alphabetically or by category. Many separately installed password managers have robust sorting and retrieval capabilities for passwords and other types of information, and this may be feature that could be emphasized to target users for whom organization is a primary concern.

Multiple participants mentioned website guidelines on account creation pages were their main source of password knowledge, so we suggest that these guidelines could offer advice beyond password composition, including nudges to use password generators and password-management tools.

**Targeting Users of Built-In Password Managers** Given that prompts to save passwords seemed to be extremely effective for many of the users of built-in password managers that we interviewed, password-generation prompts might also be effective to nudge these users to adopt safer password practices. Chrome 69, released during the course of this interview study, introduced password generation prompts for signed-in Google users [8]. However, none of our participants mentioned awareness or use of this feature. At present (summer 2019), Chrome prompts users to generate passwords by default as long as they are signed in and have turned on password syncing, which may nudge more users to use randomly generated passwords.

We did encounter one user who was using Safari's password generator and was mostly satisfied, although she encountered problems when Safari's generator did not meet password requirements. Built-in generators will likely need to offer better compatibility with website requirements to increase usability and adoption, either by offering options to adjust length or character classes manually or by automatically conforming to website requirements.

Users of built-in password managers in our study generally did not know that there was a way to view their passwords after saving them. Chrome's latest UI does seem to make this more obvious by providing a link to passwords.google.com after a password is saved, which may improve usability.

With improved password generation tools, built-in password managers could be convenient *and* secure options for users who prefer not to install specialized tools. However, some interviewees were using multiple operating systems and/or browsers and could not rely on a single ecosystem like Apple's or Google's for password storage, and these users might adopt separately installed tools if they had sufficient information and confidence in the usability of those tools.

**Targeting Users of Separately Installed Password Managers** Current users of separately installed tools, as well as some participants who had tried and failed to use those tools, often portrayed the setup process as daunting. Many current users did not update all of their accounts to have randomly generated passwords at the time of adoption, but continued to use many reused and/or weak passwords, changing them gradually over time. Some password managers offer tools that attempt to replace weak passwords automatically, but none of our participants mentioned awareness or use of such features. Some password managers, including 1Password, intentionally do not offer such features [29], but if users are thus retaining large numbers of weak or reused passwords, password-manager makers may need to offer a feature to assist with improving existing passwords at the time of adoption.

Participants who were open to using separately installed password managers did not specifically report being deterred by cost. However, when we inquired regarding cost, most participants who were not currently using separately installed password managers were unwilling to pay for password management. These participants might be more likely to try built-in password managers or password managers with free versions, such as LastPass. Some said they might be willing to pay only if the tool was "very secure" or very usable, or if it offered special features such as identity theft protection. Most who were willing to pay for password management indicated that they would pay five dollars per month or less.

## 6 Conclusion

Our analysis of 30 interviews with non-users of password managers, users of built-in password managers, and users of separately installed password managers, confirms convenience, usability, and security concerns observed in past studies of password manager adoption. We highlight barriers not previously identified, such as confusion about the source of password prompts or the meaning of "remember me" options.

We find that users of built-in password managers are often driven by convenience, whereas users of separately installed password managers prioritize security, which may explain past findings showing higher levels of password reuse among users of built-in password managers. We call for tailored designs for these two mentalities. Future work should focus on ways to serve users whose primary task is not security and nudge them to use password generators without sacrificing convenience. Our results regarding user-interface frustrations also call for better usability testing and design for password managers, including more focus on non-expert users, as well as long-term field studies to reveal edge cases in which password managers may not function as intended.

## Acknowledgments

This research was supported in part by the North Atlantic Treaty Organization (NATO) through Carnegie Mellon CyLab. We would like to thank Chelse Swoopes and Soraya Alli for their assistance with the study design and the interviews.

## References

[1] Nora Alkaldi and Karen Renaud. Why do people adopt, or reject, smartphone password managers? In *Proceedings of the 1st European Workshop on Usable Security (EuroUSEC '16)*, 2016.

[2] Nora Alkaldi, Karen Renaud, and Lewis Mackenzie. Encouraging password manager adoption by meeting adopter self-determination needs. In *Proceedings of the 52nd Hawaii International Conference on System Sciences (HICSS '19)*, 2019.

[3] Salvatore Aurigemma, Thomas Mattson, and Lori N. K. Leonard. So much promise, so little use: What is stopping home end-users from using password manager applications? In *Proceedings of the 50th Hawaii International Conference on System Sciences (HICSS '17)*, 2017.

[4] Andrey Belenko and Dmitry Sklyarov. "Secure password managers" and "military-grade encryption" on smartphones: Oh, really? Technical Report MSU-CSE-06-2, Elcomsoft Co. Ltd., https://www.elcomsoft.com/WP/BH-EU-2012-WP.pdf, 2012.

[5] Andrew Chaikivsky. Everything you need to know about password managers. *Consumer Reports*, https://www.consumerreports.org/digital-security/everything-you-need-to-know-about-password-managers, February 2017.

[6] Sonia Chiasson, P.C. van Oorschot, and Robert Biddle. A usability study and critique of two password managers. In *Proceedings of the 15th USENIX Security Symposium*, 2006.

[7] Catalin Cimpanu. Crooks reused passwords on the dark web, so Dutch police hijacked their accounts. *BleepingComputer*, https://www.bleepingcomputer.com/news/security/crooks-reused-passwords-on-the-dark-web-so-dutch-police-hijacked-their-accounts, July 2017.

[8] Catalin Cimpanu. Chrome 69 released with new UI and random password generator. *ZDNet*, https://www.zdnet.com/article/chrome-69-released-with-new-ui-and-random-password-generator, September 2018.

[9] Anupam Das, Joseph Bonneau, Matthew Caesar, Nikita Borisov, and XiaoFeng Wang. The tangled web of password reuse. In *Proceedings of the 2014 Network and Distributed System Security Symposium (NDSS '14)*, 2014.

[10] Michael Fagan, Yusuf Albayram, Mohammad Maifi Hasan Khan, and Ross Buck. An investigation into users' considerations towards using password managers. *Human-centric Computing and Information Sciences*, 7(1), December 2017.

[11] Dinei Florêncio and Cormac Herley. A large-scale study of password habits. In *Proceedings of the International World Wide Web Conference (WWW)*, May 2007.

[12] Geoffrey Fowler. Password managers have a security flaw. but you should still use one. *The Washington Post*, https://www.washingtonpost.com/technology/2019/02/19/password-managers-have-security-flaw-you-should-still-use-one, February 2019.

[13] Xianyi Gao, Yulong Yang, Can Liu, Christos Mitropoulos, Janne Lindqvist, and Antti Oulasvirta. Forgetting of passwords: Ecological theory and data. In *Proceedings of the 27th USENIX Security Symposium*, August 2018.

[14] Paolo Gasti and Kasper B. Rasmussen. On the security of password manager database formats. In *Proceedings of the 17th European Symposium on Research in Computer Security (ESORICS '12)*, 2012.

[15] Shirley Gaw and Edward W. Felten. Password management strategies for online accounts. In *Proceedings of the 2nd Symposium on Usable Privacy and Security (SOUPS '06)*, 2006.

[16] Amber Gott. Lastpass reveals 8 truths about passwords in the new password exposé. *The LastPass Blog*, https://blog.lastpass.com/2017/11/lastpass-reveals-8-truths-about-passwords-in-the-new-password-expose.html, November 2017.

[17] Joshua Gray, Virginia N. L. Franqueira, and Yijun Yu. Forensically-sound analysis of security risks of using local password managers. In *24th IEEE International Requirements Engineering Conference*, September 2016.

[18] Alex Hern. Do we really want to keep all our digital eggs in one basket? *The Guardian*, https://www.theguardian.com/technology/2015/jun/17/do-we-really-want-to-keep-all-our-digital-eggs-in-one-basket, June 2015.

[19] Troy Hunt. The only secure password is the one you can't remember. *troyhunt.com*, https://www.troyhunt.com/only-secure-password-is-one-you-cant, March 2011.

[20] Troy Hunt. Passwords evolved: Authentication guidance for the modern era. *troyhunt.com*, https://www.troyhunt.com/passwords-evolved-authentication-guidance-for-the-modern-era, July 2017.

[21] I. Ion, R. Reeder, and S. Consolvo. "...No One Can Hack My Mind": Comparing expert and non-expert security practices. In *Proceedings of the 11th Symposium on Usable Privacy and Security (SOUPS'15)*, July 2015.

[22] Blake Ives, Kenneth R. Walsh, and Helmut Schneider. The domino effect of password reuse. *Communications of the ACM*, 47(4):75–78, April 2004.

[23] Ambarish Karole, Nitesh Saxena, and Nicolas Christin. A comparative usability evaluation of traditional password managers. In Kyung-Hyune Rhee and DaeHun Nyang, editors, *Proceedings of the 13th International Conference on Information Security and Cryptology (ICISC '10)*, Seoul, Korea, 2010.

[24] J. Richard Landis and Gary G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174, 1977.

[25] Zhiwei Li, Warren He, Devdatta Akhawa, and Dawn Song. The emperor's new password manager: Security analysis of web-based password managers. In *Proceedings of the 23rd USENIX Security Symposium*, August 2014.

[26] Sanam Ghorbani Lyastani, Michael Schilling, Sascha Fahl, Michael Backes, and Sven Bugiel. Better managed than memorized? Studying the impact of managers on password strength and reuse. In *Proceedings of the 27th USENIX Security Symposium*, 2018.

[27] William Melicher, Michelle L. Mazurek, Darya Kurilova, Sean M. Segreti, Pranshu Kalvani, Richard Shay, Blase Ur, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. Usability and security of text passwords on mobile devices. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*, 2016.

[28] ThreatMetrix Digital Identity Network. Cybercrime report 2017: A year in review. *ThreatMetrix,* https://www.threatmetrix.com/info/2017-cybercrime-year-in-review, January 2018.

[29] Lars Olsson. Automatic password changing. *1Password Forum,* https://discussions.agilebits.com/discussion/87083/automatic-password-changing, March 2018.

[30] Danny Palmer. This sneaky botnet shows why you really, really shouldn't use the same password for everything. *ZDNet,* https://www.zdnet.com/article/this-sneaky-botnet-shows-why-you-really-really-shouldnt-use-the-same-password-for-everything, May 2016.

[31] Sarah Pearman, Jeremy Thomas, Pardis Emami Naeini, Hana Habib, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, Serge Egelman, and Alain Forget. Let's go in for a closer look: Observing passwords in their natural habitat. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS '17)*, 2017.

[32] Jamie Phelps. Does 1Password autofill input fields? *1Password Forum,* https://discussions.agilebits.com/discussion/62706/does-1password-autofill-input-fields, April 2016.

[33] Ellie Powers and Chris Beckmann. Chrome's turning 10, here's what's new. *Google Blog,* https://www.blog.google/products/chrome/chromes-turning-10-heres-whats-new, September 2018.

[34] Steve Ragan. Mozilla's bug tracking portal compromised, reused passwords to blame. *CSO,* https://www.csoonline.com/article/2980758/data-breach/mozillas-bug-tracking-portal-compromised-reused-passwords-to-blame.html, September 2015.

[35] Dominik Reichl. Setup - KeePass. *KeePass Password Safe* (official website), https://keepass.info/help/v2/setup.html, 2019.

[36] Richard Shay, Lorrie Faith Cranor, Saranga Komanduri, Adam L. Durity, Phillip (Seyoung) Huh, Michelle L. Mazurek, Sean M. Segreti, Blase Ur, Lujo Bauer, and Nicolas Christin. Can long passwords be secure and usable? In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*, 2014.

[37] David Silver, Suman Jana, Dan Boneh, Eric Chen, and Collin Jackson. Password managers: Attacks and defenses. In *Proceedings of the 23rd USENIX Security Symposium*, 2014.

[38] Aaron Smith. Americans and cybersecurity. Pew Research Center, http://www.pewinternet.org/2017/01/26/2-password-management-and-mobile-security, January 2017.

[39] Elizabeth Stobert and Robert Biddle. The password life cycle: User behaviour in managing passwords. In *Proceedings of the 10th Symposium On Usable Privacy and Security (SOUPS'14)*, July 2014.

[40] Elizabeth Stobert and Robert Biddle. A password manager that doesn't remember passwords. In *Proceedings of the 2014 New Security Paradigms Workshop (NSPW)*, 2014.

[41] Elizabeth Stobert and Robert Biddle. Expert password management. In Frank Stajano, Stig F. Mjølsnes, Graeme Jenkinson, and Per Thorsheim, editors, *Technology and Practice of Passwords*, volume 9551, pages 3–20. Springer International Publishing, 2016.

[42] Blase Ur, Fumiko Noma, Jonathan Bees, Sean M. Segreti, Richard Shay, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. "I Added '!' at the End to Make It Secure": Observing password creation in the lab. In *Proceedings of the 11th Symposium on Usable Privacy and Security (SOUPS'15)*, 2015.

[43] Chun Wang, Steve T.K. Jan, Hang Hu, Douglas Bossart, and Gang Wang. The next domino to fall: Empirical analysis of user passwords across online services. In *Proceedings of the Eighth ACM Conference on Data and Application Security and Privacy (CODASPY '18)*, 2018.

[44] Rick Wash, Emilee Rader, Ruthie Berman, and Zac Wellmer. Understanding password choices: How frequently entered passwords are re-used across websites. In *Proceedings of the 12th USENIX Conference on Usable Privacy and Security (SOUPS '16)*, 2016.

[45] Jeff Yan, Alan Blackwell, Ross Anderson, and Alasdair Grant. Password memorability and security: Empirical results. *IEEE Security and Privacy*, 2(5):25–31, September 2004.

# A Appendix: Open Data

The codebook, a more detailed demographic summary, and the (anonymized) dataset for this paper are available at https://osf.io/6u7m8/.

# B Appendix: Interview Script

## B.1 General Questions about Passwords

1. What types of online accounts do you have? (e.g. social media, bank accounts, shopping sites, etc.)
2. What level of protection do you think they each need? (Follow up, if necessary): Are there some accounts you want to protect more than others?
3. To the best of your knowledge, approximately how many online accounts do you have that use passwords?
4. How many of these do you access on a daily basis?
5. On which device(s) do you access these online accounts? (Follow-up below for each category the person has.)
   (a) For phones/tablets: what type(s)? (iPhone, Android, etc.)
   (b) For computers: what operating system(s)? (Windows, Mac, Linux, ChromeOS, etc.)
   (c) Public, work or personal device?
   (d) For each device: what web browser do you use most often on your [device]?
6. How many times do you manually type in passwords on a daily basis?
   (a) Which types of accounts?
   (b) On which device(s)?
7. How many of your accounts are always logged in?
   (a) Which types of accounts?
   (b) On which device(s)?
8. Do you have any passwords that get auto-filled for you?
   (a) Which types of accounts?
   (b) On which devices?
   (c) Do you know how your passwords are auto-filled?
9. Are your passwords different for each account?
   (a) (If yes) Are your passwords similar to one another?
   (b) (if reuse exists): How many of your accounts share the same password? How many of your accounts have unique passwords?
10. How do you create a password for a new account?
    (a) How does this password compare to other passwords? (i.e. is it similar?)
    (b) What if your password does not meet the character/length requirements. How would you change your password to meet those requirements?
    (c) Is this process different for some types of accounts? Which ones? What do you do?
11. How do you keep track of your passwords now? Do you use more than one method?
12. Are you satisfied with your current method(s) of managing your passwords?
    (a) What do you find easy about it?
    (b) What do you find difficult about it?
13. Has anyone ever logged into any of your accounts without your permission?
    (a) (if yes) Was this done by someone you didn't know?

(b) (if yes) What did you do? Follow up, if applicable:
   i. Did you change the compromised password?
   ii. How did you choose the new password?
   iii. How does the new password compare to your existing passwords?
   iv. Did you change the passwords to your other accounts that share the same password?

(c) (if no) What would you do if someone did?
   i. Would you change the compromised password?
   ii. (If yes) How would you choose the new password?)
   iii. How would you choose it?

14. To your knowledge, have any of your accounts ever been subject to a password data breach?

(a) (if yes)
   i. How did you find out about it?
   ii. What did you do?
   iii. After the breach, did you change the way you manage your passwords?
   iv. Did that account share a password with any of your other accounts?
   v. If so, did you change any of those passwords?

(b) (if no) What would you do if it was?

## B.2 General Questions about Password Managers

1. Have you ever heard of password managers? Where did you hear about them?
2. Do you use a password manager?
3. What, to your knowledge, is the purpose of a password manager? (If they respond to something along the lines of "it manages passwords") What else do you think they're used for?
4. *Read description of password managers to participant*

   Password managers are tools that can securely handle passwords for you. They can remember your passwords, generate new ones, and even sync them across devices. There are various types of password managers with different features, but for the purpose of this interview, we will consider three of them.

   One type of password manager is built into the web browser, such as Google Chrome, Mozilla Firefox, Safari, Internet Explorer, and Microsoft Edge. These browsers can remember passwords for websites, as well as autofill them for you.

   Another type of password manager is a third-party application. This can be software you install directly onto your devices or a service you can access on the web. It can also remember and/or autofill your passwords, including across browsers and devices.

   Lastly, your operating system can serve as a password manager as well. For example, the Keychain functionality on MacOS can remember passwords in and out of your browser. It can also be used with iCloud to sync passwords across Apple devices.

   Ultimately, the main purpose of password managers is to automatically handle your passwords for you.
5. Based on our description, which of these categories of password managers do you currently use, if any?
6. Have you used any [other] password manager tools in the past?
7. (If they have used PM, now or in the past) When did you start using a password manager? Why did you start using it?
8. (if stopped use): When did you stop using the password manager and why?
9. (If they use any and haven't already named them) Can you name the password management tools that you use? (Or if they can't name them, ask them to describe them / indicate how they use them so that you can try to discern what they mean)

## B.3 Experience Using Password Managers

1. Why did you choose [PM]?
2. How has your experience been using a password manager?
3. What functions did you like / find helpful?
4. What functions did you dislike / find unhelpful?
5. Is all functionality of your password manager available for free, or does this tool have a paid version?

   (a) (If paid version exists) Do you use the paid or free version? Why?

   (b) (if uses free version) Would you ever pay for a password manager? How much? What features would it have?
6. Do you use your password manager on all of your devices, including [list of tools they already told you about in the first section]?

   (a) (if no)

i. Which devices do you use it on?
ii. Why do you use it on those?
iii. Why not use it on the others?
iv. How do you keep track of passwords on the device(s) that you don't use your PM on?

7. (For each device that the user uses PM on): Did you have to install an application to your device, or install an extension to your browser, or both?
   (a) (if no) How do you access your password manager? (possible answers include logging into a website, or USB drive)
8. Does your password manager offer the option of syncing passwords between devices?
   (a) (If this option exists) Do you use it? Why or why not?
9. Do you use your password manager for all the accounts you access through your web browser?
   (a) If not, how do you decide which accounts to use it for?
   (b) How do you keep track of passwords that are not stored in this PM?
10. Do you use your password manager for any accounts outside of your web browser? Examples of this would include an email client like Outlook on your computer or a social media app such as Facebook on your phone.
   (a) Do you use it for all of the accounts outside of your web browser(s)?
   (b) (if no to a) How do you decide which accounts to use it for?
11. Do you have to provide a master password or other authentication to access the passwords stored in your password manager?
   (a) (If yes) What type of password or authentication is required?
      i. (if master password):
         A. How did you create your master password?
         B. Is your master password similar to your other passwords?
         C. Is it difficult to remember your master password?
         D. (if yes) How do you remember it?
   (b) (If yes) How often do you have to provide it?
   (c) Have you ever modified the default settings to change how often you have to provide this?
12. Do you feel like your passwords are safe and secure when stored in this PM tool?
13. Do you know how this tool protects the security of your passwords? *(Unless they say they have no idea, ask them to elaborate on how they think it works)*
14. Does your password manager have a password generation tool?
   (a) (if yes to 14) Have you ever used the password generation tool?
      i. (if yes to a):
         A. Do you use the generation tool for newly created accounts?
         B. Have you used the tool to generate a new password for an existing account?
         C. (if yes to B) Does your password manager have an automatic password replacement feature that changes passwords for you without you having to actually visit the website yourself? Do you use it? Why or why not?
         D. Approximately how many of your passwords are now created by the password generation functionality?
         E. Do you ever change the settings from the defaults when generating a password?
         F. Was there an instance where the generated password did not meet the website's password requirements? (If yes) What did you do about it?
         G. Overall, how has your experience been using the password generation tool?
15. Does your password manager have a dashboard or tool that examines the security of your passwords?
   (a) (If yes): How often do you use it?
   (b) Have you changed any of your passwords after looking at this information?
16. Has your password manager ever informed you of a data breach? (If yes) What did you do?
17. Has your password manager ever prompted you to change your password? (if yes) Under what situation? What did you do?
18. Are there any additional services or features that you would want in your password manager?

## B.4 Why not Using PMs? (If answer "no" to Using Password Managers)

1. Can you tell us why you aren't using a password manager? *(Follow up by probing what it would take for them to use a password manager.)*
2. Many third-party password managers require a monthly fee to use their services. Would you be willing to pay for such a service?
   (a) If yes, how much?
   (b) If no, why not? (*If participant says there are free third-party PMs available, then ask:* Would you be willing to pay for additional features that are not included in the free version? How much would you be willing to pay?)

## B.5 Perceptions of Password Managers' Functions

We talked about different types of password managers a few minutes ago, including third-party password managers, password managers built into web browsers, and password managers built into operating systems.

1. Do you think some types of password manager tools are safer to use than others? (Why?)
2. Do you think some types of password manager tools are more convenient than others? (Why?)
3. How do you think password manager tools compare to other methods of managing passwords, such as writing them down on paper or saving them in a file on your computer? (Why?)
4. How do you think password managers store passwords?
   (a) Do you think password managers store your passwords locally on your device or on a server (in the cloud)?
      i. Do you think one is more secure than the other? (If so, which one? Why?)
      ii. Do you have a preference? Why or why not?
   (b) How do you think password managers sync your accounts across devices?
      i. Would you want this function? Why?
      ii. Do you think this impacts your password security? If so, how?
   (c) What do you think the password data looks like when stored on your computer?
      i. If your password is "password2018!", does your password manager store it as "password2018!"?
      ii. Is there a difference when stored in the cloud?
5. Do you think password managers affect the security of your accounts? Why or why not?
6. Do you trust password managers to always store or not forget your passwords? Why or why not?
7. Do you trust password managers to protect your passwords from attackers? Why or why not?
8. Have you ever received advice or training on how to create or manage passwords?
   (a) (if yes) What guidelines have you been taught? Where?
   (b) (if yes) Do you use these guidelines? Why or why not?
9. (non-PM user): Would you consider using a password manager in the future? Why or why not?
10. (If "No" or "I don't know" to data breach question in Part B.1, Q. 13: Earlier you mentioned that you were never impacted by a data breach, or that you weren't sure if you were. Would you like the opportunity to verify this? We can use a website called HaveIBeenPwned to check whether your accounts were compromised in a public data breach.
    (a) *Explain to participant:* The website asks for your email address and checks if any accounts tied to it were compromised. Note, however, that the website cannot check information on every data breach. It checks those that are known to the public.
    (b) *If participant agrees, inform participant:* For privacy reasons, we recommend that you access the website on your own device. This way, we won't see your email address, nor will we know which of your accounts, if any, were impacted by a breach.
    (c) *Instruct the participant to try any other email address they may use often.*
    (d) Were any of your accounts compromised?
    (e) (if yes)
       i. How many?
       ii. What types of accounts? (Provide categories to choose from: social media, bank, shopping, other)
       iii. How do you feel about this information?
       iv. (follow up, if necessary) Will you do anything with this information?

## C Appendix: Codebook

Code categories are shown in bold type, with the list of codes in that category following. For a more detailed version with code descriptions, see https://osf.io/6u7m8/.

- **Account type:** shopping, banking, utilities, email, social media, healthcare, work, school, other
- **Account number:** less than 10, 10-15, 16-20, 21-30, 31-50, 51-100, more than 100
- **Account importance:** all accounts important, financial accounts, accounts with PII, work accounts, Facebook, email, other
- **Accounts accessed daily**: accounts accessed daily *(single code to used only identify snippet where participant gave estimate of this number)*
- **Password composition:** use passwords of equal strength, stronger passwords for more important accounts, disposable/weaker passwords for unimportant accounts, unique passwords for all accounts, unique passwords for important accounts, use shared substrings, use randomly generated, use passphrase, use words related to website type, use 2FA for more important accounts
- **Devices and browsers used:** iPhone/Safari, iPhone/Chrome, iPhone/Firefox, iPhone/other, iPad/Safari, iPad/Chrome, iPad/Firefox, Windows/Chrome, Windows/Firefox, Windows/Edge, Windows/IE, Mac/Safari, Mac/Chrome, Mac/Firefox, Android/Chrome, Android/Firefox, Android/other, Linux/Chrome, Linux/Firefox, Linux/other
- **Passwords typed daily:** 0, 1-2, 5, other number
- **Passwords saved:** never, unimportant accounts,
- **Exceptions to password reuse:** set by someone else, need to share, use old password, forced change, password requirements, other
- **Exceptions to password reuse: method of remembering exception password:** write down, other
- **Action when password is rejected due to password creation requirement:** add required characters, regenerate new password, remove forbidden characters, other
- **Password creation process:** same password for all accounts, use generator, one password per "tier" of accounts, use memorable personal info, other
- **Current password management:** synced file, guessing variations / resetting, physical notes, local file, memory, third-party PM, keychain, browser, fingerprint, not sure, other
- **Password management: satisfied?:** satisfied, not satisfied, not sure
- **Password management likes (non-PM methods):** always accessible locally, easy to remember, other
- **Password management dislikes (non-PM methods):** potential to lose, hard to remember, other
- **Had compromised account:** yes, no, I don't know
- **Compromised account action:** major/total change to compromised password, minimal change to compromised password, contact support, change passwords for accounts with same email, stronger password, other
- **Had data breach:** yes, no, I don't know
- **Data breach action:** change password, contact support, change passwords for accounts with same email, other
- **Aware of password managers?:** aware, not aware
- **Not use PM reason:** not many accounts, not aware of PMs, not much to protect, security concerns, master password concerns, past negative experience, other
- **Heard of PM from:** work, media, other people, I don't know, other
- **PM definition:** store/organize passwords, unique passwords, generate random passwords, no need to memorize, improve security, autofill, I don't know, other
- **Use PM time:** less than 1 year, about 1 year, multiple years
- **Use PM device:** all, non-shared, computers only *not phones, tablets, etc.*
- **Start using PM reason:** convenience, memory limitations, receive prompts, security, other
- **PM like function:** autofill, generate strong passwords, no memorizing, syncing, unique passwords, view passwords, desktop client, other
- **PM dislike reason:** incompatible device, saved unwanted passwords, cannot view passwords, generates passwords with unacceptable symbols, other
- **PM feature request:** PM feature request *(single code used to tag all snippets referring to features that participants wished PMs had)*
- **PM switch strategy:** gradually, change all at start
- **Use PM to store info other than website passwords:** use PM for application passwords, use PM for other info (e.g. credit cards)
- **Master password unique:** yes, no
- **Master password composition:** random, passphrase, other
- **Uses 2FA in combination with master password**: yes *(single code only used for participants who reported using this combination)*
- **Pays for PM (if using) or willing to pay (if not using a PM)?:** yes (currently pays or would pay), no (does not pay / would not pay), depends
- **Function that would convince them to pay for PM**: 2FA *(single code, no other specific functions mentioned)*
- **Not pay for PM reason:** already using free version, other
- **Pay for PM price:** $5 or less per month, depends, other
- **PM dashboard:** has used, has not used, not available in their PM (as self-reported)
- **Exceptions, PM users: certain passwords not stored in PM:** infrequently used, habit, multiple accounts, personal info, shared computer, email, financial, old account
- **PM generator:** not aware, aware / does not use, aware

/ does not use now / would not unless something "bad" happened, uses, not available for their PM (as self-reported)
- **Choosing PM reason:** compatibility, cost/value, features, convenience, other
- **PM security (beliefs about most secure type):** OS most, third-party most, browser most, third-party least, browser least, depends, no difference, I don't know
- **PM security belief reason:** not sure, connected to internet → less secure, trusts known names (e.g., Google), distrusts third parties, password storage method (e.g., "browser not secure because it stores passwords in plaintext"), trusts specialized/password-specific tools, distrusts browser code, other
- **PM convenience (opinions about most convenient type):** third-party least, browser most, OS most, third-party most, no difference, I don't know
- **PM convenience belief reason:** no extra setup step, no extra cost, not sure, other
- **PM or other methods safer?:** PM, other, neither/unclear/depends
- **Beliefs on where PMs store passwords: locally or cloud?:** locally, cloud, both, depends, I don't know
- **More secure: locally or cloud?:** locally safer, cloud safer, equal, I don't know
- **PM stores passwords: format?:** "in code," encryption, plaintext, I don't know
- **PM stores passwords: format: different in cloud?:** no difference, difference, I don't know
- **PM effect on security:** no effect, positive effect, negative effect, I don't know
- **Trust PM to remember passwords?:** sometimes, yes, no
- **Trust PM to protect from attackers?:** not sure, yes, no
- **Password advice sources:** website guidelines, people, work, other
- **HaveIBeenPwned:** used previously, no pwnage found, breach found, declined
- **Uses syncing?:** currently syncs, does not currently sync
- **Syncing perceptions?:** useful, not useful secure, not secure
- **Miscellaneous themes:** acknowledges risks, uses 2FA for all accounts, dormant accounts, time-saving, does not log out of accounts, social media: no customer support, conflates hacking and data breach, stores work passwords in PM, work passwords similar to personal, uses PM at work, logs out of banking accounts, likes having constant/mobile access to passwords, access problems, control, attackers target bigger entities than me, inertia b/c no past bad experiences, confusion about "remember me", specifically trusts Google, specifically trusts Apple, specifically trusts another company, tradeoffs, frustration with security questions, account sharing, knowledge gap, device sharing, habit, self-imposed password changes, kept reused passwords after breach, avoids creating new accounts, multiple accounts on same website, bank will reimburse, switching too much work, browser not safe, threat model: physical access, negative past PM experience, required password changes at work, all eggs in one basket

# Of Two Minds about Two-Factor: Understanding Everyday FIDO U2F Usability through Device Comparison and Experience Sampling

Stéphane Ciolino
*OneSpan Innovation Centre & University College London*
*stephane.ciolino.17@ucl.ac.uk*

Simon Parkin
*University College London*
*s.parkin@ucl.ac.uk*

Paul Dunphy
*OneSpan Innovation Centre*
*paul.dunphy@onespan.com*

## Abstract

Security keys are phishing-resistant two-factor authentication (2FA) tokens based upon the FIDO Universal 2nd Factor (U2F) standard. Prior research on security keys has revealed intuitive usability concerns, but there are open challenges to better understand user experiences with heterogeneous devices and to determine an optimal user experience for everyday Web browsing. In this paper we contribute to the growing usable security literature on security keys through two user studies: (i) a lab-based study evaluating the first-time user experience of a cross-vendor set of security keys and SMS-based one-time passcodes; (ii) a diary study, where we collected 643 entries detailing how participants accessed accounts and experienced one particular security key over the period of one week. In the former we discovered that user sentiment towards SMS codes was typically higher than for security keys generally. In the latter we discovered that only 28% of accesses to security key-enabled online accounts actually involved a button press on a security key. Our findings confirm prior work that reports user uncertainty about the benefits of security keys and their security purpose. We conclude that this can be partly explained by experience with online services that support security keys, but may nudge users away from regular use of those security keys.

## 1 Introduction

User authentication mechanisms are based on one of three factors: *knowledge* (e.g., password), *ownership* (e.g., token), or *inherence* (e.g., fingerprint). Combining factors (e.g., passwords and tokens) is widely recognized as an effective technique to protect both corporate and personal online accounts against account hijacking threats. Indeed, there are already examples of citizens being advised to use Two-Factor Authentication (2FA) by government agencies (as in the UK [35]). The most common second factor is a One-Time Passcode (OTP) received via a text message to a mobile device [5]. While this technique conveniently leverages pre-existing telecommunications infrastructure and mobile devices that users already own, it is vulnerable to Person-In-The-Middle (PITM) attacks [8, 23, 32], such as phishing attacks. Hardware-based authentication tokens have historically been deployed for 2FA in closed user communities such as corporations, or for banking customers, particularly in territories such as Europe.

Recently, efforts have gathered pace to position tokens as a general purpose second factor for end-users to secure a range of online accounts. The Fast IDentity Online (FIDO) Alliance was founded in 2012 to reduce reliance on passwords for Web authentication by moving to new authentication standards underpinned by public key cryptography that are resistant to phishing [2]. *Security keys* are commercially available authentication tokens based on the Universal 2nd Factor (U2F) standard created by FIDO [3]. They are currently supported by more than 30 Web service providers [47] including Dropbox, Facebook, GitHub, Google, and Twitter.

However, there are barriers to the widespread uptake of security keys for 2FA. These include the need for an improved setup process [16], and inherent concerns about losing the devices [1, 38]. In the broader debate, there are mixed views about the decisive factors that might influence the uptake of security keys for everyday use. Research has shown that some users struggle to see the utility in security keys [16], yet other work reports user satisfaction with the devices [38].

In this paper, we aim to further understand user perceptions of utility and security of the security keys. We firstly compared the setup and login experience for three cross-vendor security keys: Feitian ePass FIDO® NFC (ePass), OneSpan DIGIPASS SecureClick (SecureClick), Yubico YubiKey 4 Nano (YubiKey); and SMS-based OTPs. Participants used


*USENIX Symposium on Usable Privacy and Security (SOUPS)* 2019.
August 11–13, 2019, Santa Clara, CA, USA.

each mechanism on two representative Web services: Gmail and Dropbox. We discovered that the security keys generated diverse usability issues and that the Web service user interface could also impact the efficiency of the setup process. We built on the lab-based study with a week-long diary study of one specific security key (the SecureClick) involving fifteen participants, each using a SecureClick with free choice of their online accounts. We collected 643 diary entries and found that participants only used security keys in 28% of logins that *could have* used a security key. Also, we found that button presses on the security key decreased by 50% from the first day of the study to the last.

The rest of the paper is arranged as follows. Related Work is discussed in Section 2. We describe the protocol and results of our laboratory study in Section 3. The protocol and results for the diary study are described in Section 4. We close the paper with Discussion and Conclusion (Sections 5 and 6 respectively).

## 2 Related Work

### 2.1 2FA Mechanisms

Research is increasingly exploring the usability and security of tokens that provide 2FA based on a pre-shared secret between the token and the service provider. De Cristofaro et al. [18] surveyed the usability of various forms of 2FA: codes generated by dedicated tokens or smartphone apps (e.g., Google Authenticator), or received via email or SMS. Respondents' perception of 2FA usability correlated with distinguishing characteristics of each mechanism; the actual 2FA technology deployed or the context of use had less of an influence. Three metrics were argued to be necessary for rating the usability of 2FA technologies: ease-of-use; required cognitive effort; and trustworthiness of the device.

Other studies have examined 2FA in the context of online banking (e.g., [7, 30, 44, 45]). Weir et al. [44] studied three different tokens used by banking customers and found that participants preferred those with higher perceived convenience and usability, at the expense of perceived security. A follow-up study [45] contrasted password-based authentication against token- and SMS-based 2FA, finding that convenience, personal ownership, and prior experience were key factors in selecting an authentication mechanism. Krol et al. [30] report that the mental and physical workload required to use tokens influenced user strategies for accessing online banking (e.g., how often they would be willing to log in). Althobaiti and Mayhew [7] conducted an online survey across students studying abroad, identifying higher perceived usability for tokens over SMS-based authentication.

Weidman and Grossklags [43] examined the transition from an authentication token to a 2FA system using DuoMobile on employees' personal mobile devices within an academic institution. Users rated the DuoMobile solution more negatively compared to the token-based solution, as users resented using their personal mobile devices in a work context.

Gunson et al. [25] recruited banking customers to contrast knowledge-based one-factor authentication (1FA) and token-based 2FA for automated telephone banking. No single 1FA or 2FA approach stood out as a preferred authentication method. However, a trade-off between usability and security was identified, with 1FA judged more usable but less secure than 2FA. Sasse et al. [40, 42] examined authentication events involving passwords and RSA tokens in a US governmental organization – authentication disruptions reduced staff productivity and morale, to the extent that work tasks were arranged to minimize the need to authenticate.

### 2.2 FIDO U2F and 2FA Security Keys

Lang et al. [32] applied the usability framework established by Bonneau et al. [11] to assess a range of security keys, alongside authentication activity data from Google. The authors identified that security keys evidenced quicker authentication and fewer support incidents in a work environment, as compared to alternative tokens. Das et al. [16, 17] conducted a two-phase laboratory study with students, to improve the usability of setting up a Yubico security key with Gmail. Participants did not perceive benefits to using the Yubico security key in their everyday lives and were most concerned with the potential of losing access to their account. Colnago et al. [15] examined the adoption of Duo 2FA at a university. Security keys were one of four 2FA options offered to users, with less than 1% choosing this option. Reynolds et al. [38] explored usability perceptions of Yubico security keys during enrolment, and in everyday use (by way of a diary study). They found that participants experienced problems to set up the security keys with services but perceived them as usable for regular activities. As also found in the work by Das et al. [16, 17], losing the security key was also highlighted as a concern.

### 2.3 Open Challenges

Authentication tokens have been prevalent for many years in closed and centralized deployments, e.g., in workplaces or for individual banking services in some countries. These represent service-centric technologies which are centralized and orchestrated by the service provider. Security keys are intended to support *user-centricity* [10] which is a term that has specific connotations in digital identity of: decentralization, user control, selective disclosure, and interoperability. With security keys, this user-centricity is achieved through public key cryptography: the security key can generate private keys that are stored confidentially on the device and can create digital signatures that may be shared with a service provider to attest to ownership of a given public key.

Adoption of security keys has eradicated account takeover at Google [28]. However, user adoption of security keys more generally is low, with evidence that 1% of observed logins across one entire user population leveraged security keys [21], and 1% of users in a sample at a university were using these devices [15]. Furthermore, there are strong technical arguments against the use of other more popular 2FA mechanisms today due to the threat of person-in-the-middle-attack, particularly SMS-based OTPs [6, 27, 48]). The threat against SMS-based OTPs has taken on a new dimension in recent times due to the emerging prevalence of mobile device SIM swap attacks [46].

Research up to now has been valuable to provide an early understanding how the form and function of security keys themselves impact adoption, but it has limitations: the lab work of Das et al. [16, 17] wholly focused on the setup of one YubiKey with Gmail; Reynolds et al. [38] conducted a between-subject lab study for device setup with one YubiKey and a diary study limited to Gmail, Facebook and Windows 10 with a YubiKey that did not capture specific login events; the main insight about security keys in the work of Colnago et al. [15] is that they were rarely adopted. While there is no single answer to the low adoption of security keys, we were interested in obtaining a new perspective on the user experience of security keys and compatible online services in everyday Web browsing, driven by the following research question: *Are there 2FA usability issues that security keys perpetuate, or new issues that they introduce in an everyday context, at setup, login, and service use?*

Our work, presented in the following sections, contributes to the state of the art in the following ways: (i) it compares several 2FA mechanisms with each other; (ii) the diary study accommodates free choice of Web services and captures daily interaction data; and (iii) it begins to shed light on findings from prior work why users might fail to see a benefit in the use of security keys.

## 3 Lab-based Usability Study

We conducted a lab-based study in August 2018, to capture the main factors that affect the usability and security perceptions of security keys at the setup and login phases of use.

### 3.1 Method

We conducted a within-subjects research lab-based study to compare several 2FA methods directly against each other in a way that maximizes the number of data points collected per participant. We recruited 15 participants via flyers posted on the university campus. Convenience sampling was employed, with no pre-screening applied. Each lab session lasted approximately 45 minutes and involved a series of tasks and a debrief discussion to be completed. Participants received a complimentary £10 Amazon voucher upon completing the study.

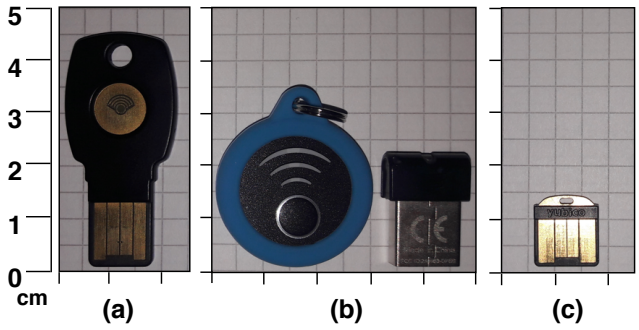



Figure 1: Visual comparison of each security used in the laboratory study: (a) ePass, (b) SecureClick, and (c) YubiKey.

The study required participants to use each of the following four 2FA mechanisms:

- ePass security key
- SecureClick security key
- YubiKey security key
- SMS-based OTPs

The security keys, shown in Figure 1, were chosen as the manufacturers (Feitian, OneSpan, and Yubico) are part of the FIDO Alliance board, and the devices themselves are diverse in form factor. The ePass and SecureClick are dual-mode devices that work with both laptops/desktops, via USB, and mobiles, via Near Field Communication (NFC) or Bluetooth Low Energy (BLE). Our study wholly focused on laptop/desktop usage. With the ePass and YubiKey, users need to plug the security key into a USB port and press the button/handle on the device to execute its functionality. The SecureClick comes in two parts, with a USB Bluetooth Bridge (as in Figure 1), requiring installation of a browser extension to link the Bluetooth Bridge and SecureClick before first use on laptops/desktops; users then only need to plug the Bluetooth Bridge into a USB port and press the button on the SecureClick itself. The study also included SMS-based OTPs since the mechanism is typically present in a 2FA choice architecture [37] competing with other methods of 2FA, and may affect users' perception or preference towards security keys.

Participants used each authentication mechanism with one of two mainstream Web services that support the above 2FA techniques: Dropbox or Gmail. We randomly assigned eight participants to use Dropbox while we assigned Gmail to the other seven participants. An earlier pilot test informed the decision to focus each participant on one of the Web services rather than both, to reduce the risk of fatigue. Participants used email accounts created solely for the study, with one per Web service (Dropbox, Gmail) and 2FA method (SMS-based OTPs, ePass, SecureClick, and YubiKey). We chose the

usernames to be easy to recall and type for the participants, and the password was the same for all of the accounts.

An abstraction of the 2FA setup and login processes for both Web services are illustrated in the Appendix, in Figure 5. This diagram relates the technical mechanisms and activities under observation [41] to user experiences and perceptions emerging from the studies (revisited in the Discussion in Section 5).

#### 3.1.1 Procedure

We followed the procedure below during the lab study:

1. **Preparation:** The participant sat at a desk in the laboratory room (the same desk for each session). The experiment moderator followed a script to explain the study to the participant. We provided an information sheet and consent form to the participant, who was then allowed time to read the forms before providing their consent. The participant would be encouraged to *think aloud* during the subsequent tasks [14].

2. **Testing Different 2FA Methods:** The main part of the laboratory session consists of four 2FA tasks. Each task has a 'set-up' phase (for enrolling a 2FA mechanism with a Web service), and a 'login' phase (using the 2FA mechanism for login with the Web service):

   - *Task A:* 2FA using SMS-based OTPs.
   - *Task B:* 2FA using ePass.
   - *Task C:* 2FA using SecureClick.
   - *Task D:* 2FA using YubiKey.

   The instructions we gave to participants are detailed in Appendix A. We varied the order of tasks A, B, C, and D across participants to minimize ordering effects on participant preference and behavior. The participants also completed a System Usability Scale (SUS) assessment of the technology immediately after each task.

3. **Debrief:** After the structured tasks, the researcher debriefed the participant in a semi-structured interview, shared the study goals, and encouraged a focused discussion. Debrief questions explored issues around 2FA, participant satisfaction/dissatisfaction with security keys, and perceptions of where security keys could be useful (or not).

#### 3.1.2 Test Equipment

Participants performed all tasks on a Dell Latitude E5540 laptop using the Windows 7 operating system, the Google Chrome browser, and a Motorola XT1100 Nexus 6 mobile phone (for SMS-based OTPs). We used a voice recorder to capture 'think-aloud' responses and the debrief dialogue, to facilitate transcription at a later stage. Interactions with the Web services were also video-recorded for timing purposes, recording only the laptop screen and page/screen transitions (not the participant).

#### 3.1.3 Research Ethics

The study was approved through the sponsor university's IRB-equivalent research ethics committee, Project ID 5336/010, and raised no specific cited concerns nor requested corrections. After we completed the study, we thoroughly debriefed participants and compensated them immediately for their time.

#### 3.1.4 Demographics

We recruited fifteen participants for our study (6 female, 9 male). The ages of the participants were between 21- and 37-years old (median 25.5). Eight were postgraduate, three were graduate, and four were up to undergraduate level. Nine already had experience of using 2FA (either SMS-based OTP or mobile-based authentication app). None had any familiarity with security keys.

#### 3.1.5 Limitations

Participants' behavior and views of the authentication mechanisms may have been shaped by the laboratory conditions (controlled to uphold internal validity [31]). Furthermore, the lab study did not present a real risk to the participants' personal data (where this is addressed in the diary study, Section 4), and required participants to use machines and accounts provided by the experimenters. Participants were comparing a relatively new authentication method (security keys) to a well-known incumbent (SMS-based OTPs), where evaluating a security technology against others in the same session has the potential to encourage more critical feedback [29]. Although our sample of 15 participants is above the recommended minimum of 12 participants to achieve data saturation [24] (achieved after 11 interviews in our case), the sample could be considered as modestly sized. We aimed to mitigate this concern in this study by capturing a detailed range of data points with our within-subjects study design and debrief interviews.

### 3.2 Results

The following sections present quantitative results (Sections 3.2.1 to 3.2.3) and qualitative results (Sections 3.3.1 to 3.3.3) pertaining to our research question.

#### 3.2.1 Phase 1: Setup

We captured critical events that prevented participants from progressing with a task (without further assistance), and present a timing analysis of setup interactions with each

| Issue | Source | Severity | Frequency |
|---|---|---|---|
| Bluetooth pairing errors on SecureClick | Device | Major | 12 |
| Generally unsure how to achieve their goal based upon available instructions | Web Service & Device | Major | 12 |
| Confusion due to SMS setup brought to the fore before mention of security key | Web Service | Major | 6 |
| Unsure whether to allow Chrome browser to see make and model of security key | Browser | Minor | 6 |
| Animated circle misinterpreted as loading by users | Web Service | Major | 5 |
| SMS-based OTP not received to set up secondary authentication | Web Service | Major | 2 |
| Unable to locate the button on a specific device | Device | Major | 2 |
| Inserting the YubiKey the wrong way up | Device | Major | 2 |

Table 1: Issues encountered when setting up 2FA technologies with a Web service, alongside their severity and frequency of occurrence. All but one issue is rated as having 'Major' severity.

security key on each Web service.

**Usability Roadblocks:** Table 1 lists the most common roadblocks that users encountered during the setup of the security keys. We use the Nielsen rating system to categorize the severity of those usability roadblocks [36]. 'Major' usability problems may cause a lot of confusion or result in the incorrect use of the system, whereas 'Minor' usability problems indicate a delay or inconvenience in the completion of a task.

Participants generally needed guidance to pair the SecureClick with the Bluetooth Bridge, citing a specific need for more clarity in the instructions and error messages displayed.

Twelve participants needed guidance to navigate the Dropbox and Gmail Web service interfaces to activate 2FA. One crucial issue was that both of the Dropbox and Gmail interfaces prioritized the process with the activation of SMS-based OTPs. The option to use a security key was not salient to participants who were unsure if they were on the correct path to activate a security key. Once participants had finally discovered the correct option (by ignoring their initial intuition), several minor user interface design issues disrupted the user journey. First, both services displayed an animated spinning circle while waiting for the user to touch the button on their security key after users inserted it into the USB port. At this point, we saw participants conclude that the website was loading rather than prompting for the button to be pressed on the security key. Five participants specifically asked for help as to whether they were required to do anything at that stage.

Another issue was that once users pressed the button on their security key, a pop-up window appeared in the browser asking the user to confirm that the Web service had permission to 'See the make and model of your security key'; six users were unsure whether to allow this since they were not forewarned that it would occur, and weren't sure if this option would breach their privacy beyond the basic use of the security key itself.

**Learnability and Efficiency:** We used the video recordings to retrospectively measure setup timings for the 2FA techniques on each online service. The measurement started when participants accessed the login page of the Web service and ended when participants viewed confirmation from the Web service that the 2FA setup was complete. The timings to set up 2FA with different mechanisms and Web services are shown in Table 2.

The median time to set up the ePass was 2min 29s and 2min 49s on Dropbox and Gmail respectively.

The median time to set up the SecureClick was 2min 23s and 2min 25s on Dropbox and Gmail respectively. Also, there was a one time process required to download the DIGIPASS SecureClick Manager and pair the SecureClick with the Bluetooth Bridge (median time was 3min 06s).

The median time to set up the YubiKey was 5min 20s and 2min 06s on Dropbox and Gmail respectively. The timings on Dropbox were impacted by device form factor and user interface issues: participants were confused about the location of the button on the YubiKey (7 participants); had to be guided through the Dropbox user interface (4); inserted the YubiKey the wrong way around (1); didn't receive the SMS-based OTP and had to restart the process (1).

The median time to set up SMS-based OTPs was 2min 33s and 1min 41s on Dropbox and Gmail respectively.

We had no a priori hypotheses about significant interactions that could emerge between the devices and services. However, we noted patterns in the data that led us to conduct post hoc analysis. A Kruskal-Wallis test uncovered significant differences in the setup times, considering the specific Web service and device as factors: $\chi^2 = 18.0366$ $p < 0.05$. Pairwise comparisons yielded significant differences between (i) YubiKey on Dropbox and YubiKey on Gmail ($p < 0.05$); (ii) YubiKey on Dropbox and SMS-based OTPs on Gmail ($p < 0.01$). The p-values included Bonferroni Correction for multiple comparisons.

#### 3.2.2 Phase 2: Login

As with the setup phase, we used the video recordings to measure 2FA login timing. We started the measurement when

| | Dropbox | | Gmail | |
|---|---|---|---|---|
| **2FA Method** | **Setup** | **Login** | **Setup** | **Login** |
| *ePass* | 149 (76) | 22 (11) | 169 (99) | 24 (11) |
| *SecureClick* | 143 (80) | 28 (20) | 145 (92) | 33 (8) |
| *YubiKey* | 320 (139) | 29 (15) | 126 (42) | 25 (7) |
| *SMS OTPs* | 153 (116) | 50 (20) | 101 (43) | 38 (21) |

Table 2: Median timings (and interquartile range) in seconds for each 2FA method and each Web service tested. Timings rounded to the nearest integer.

participants accessed the login page of the Web service and stopped once the participant had successfully logged in.

The timings for participants to perform 2FA login with different mechanisms and Web services are shown in Table 2. Excluding outliers, it seems that logging in using security keys is faster than using SMS-based OTPs. This is presumably due to the time the user must wait to receive the SMS-based OTP and then type it on the user interface.

The median time to perform 2FA login using the ePass was 22s and 24s on Dropbox and Gmail respectively.

The median time to perform 2FA login using the SecureClick was 28s and 33s on Dropbox and Gmail respectively. The timings were impacted by one participant holding the button down for too long on the SecureClick, and another waiting idle before pressing the button.

The median time to perform 2FA login using the YubiKey was 29s and 25s on Dropbox and Gmail respectively. The timings were impacted by three participants waiting some time before touching the handle on the YubiKey.

The median time to perform 2FA login using the SMS-based OTPs was 50s and 38s on Dropbox and Gmail respectively.

#### 3.2.3 2FA SUS Scoring

Participants completed an SUS rating scale after each of the four tasks. 2FA using SMS-based OTPs, the ePass and the YubiKey were all deemed 'acceptable' [9] with a mean score of 85.17 ($SD = 8.37$, 95% $CI = \pm 4.24$), 80.5 ($SD = 19.58$, 95% $CI = \pm 9.91$) and 73 ($SD = 28.16$, 95% $CI = \pm 14.25$) respectively. The SecureClick had a mean score of 61.5 ($SD = 22.93$, 95% $CI = \pm 11.60$) which is deemed 'marginal'. The distribution of the SUS scores for each mechanism are illustrated in Figure 2.

## 3.3 Qualitative Results

The data was analyzed using thematic analysis [12]. All participant responses were coded in several stages by one researcher, initially as low-level labels, moving to higher-level analytical categories. We identified seven high-level categories (with the three most prominent presented in the following sub-sections),

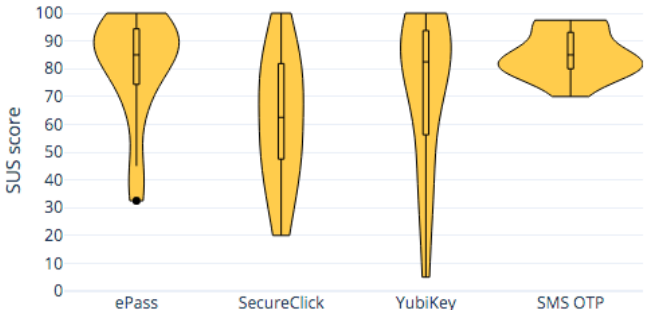



Figure 2: System Usability Scale (SUS) scores for each method of 2FA.

and 31 sublevel codes. Analysis is presented here, including notable quotes (we refer to individual participants using PL##).

#### 3.3.1 Effort, Convenience, and Fearing the Worst

Nine participants were concerned about being locked out of their accounts if using 2FA, should they not be able to present the security key or SMS-based OTP. One solution proposed to mitigate this problem was keeping a backup security key (PL06): *"so like when you get a car you get a spare key, this is something that you would need for me I think, as a backup if you lose one."*

Indeed there was an intuitive awareness that account lockout and recovery issues are easier to resolve with SMS-based OTPs. Participants also expressed being comfortable with 2FA using SMS-based OTPs, with one participant in particular (PL07) believing this to be an appropriate approach for people who are not *"tech-savvy."* Delays in receiving an SMS-based OTP, or having to swap SIM cards or enable roaming while traveling abroad, were issues voiced about SMS authentication.

It was widely recognized among participants that security keys removed the wait that is inherent to the delivery of SMS-based OTPs; PL04: *"much more efficient than codes verification and stuff like that," "it kind of removes all that leg work," "you can kind of just tap in and it's done."* However, security keys are not as versatile as mobile devices and comprise an extra object that users would have to procure, protect and carry around, as noted by PL13: *"Nowadays everyone has a phone that you carry around with you, for me to carry an extra piece which is this, it doesn't have any function other than just stick in into a computer. I mean a phone is something you need, you need it to call, it has multi-functions, it is a multi-functional thing."* The use of SMS-based OTPs (PL14) *"works well since it is using your phone, it's not something that requires an extra piece of equipment or hardware."*

One participant (PL08), who reported only using password-based 1FA, generally sees 2FA as an extra step slowing down

every single login process: *"I'm bound to using the two-step every time I want to log in, again it adds on a few extra seconds to the login process."* Others had encountered issues setting up 2FA on other online services which have created a negative perception of all 2FA procedures generally, e.g., PL09: *"I try to avoid two-step verification because I once did it as my Apple and then it got really messed up, so it's a bit hard because my phone didn't get the text so I disabled it, so just to avoid that I don't do it."*

Seven participants mentioned that having to pair the SecureClick with its Bluetooth Bridge was a drawback, e.g., PL03: *"I need to install things before I get to use it, for the moment I seem to be able in most devices."*

#### 3.3.2 Size Matters: Loss, Breakage and Design Choices

The form factor of specific devices influenced perceptions of usability. Thirteen participants commented on the unusually small size of the YubiKey, e.g., PL12: *"this one is more discrete, you can't really see it,"* with five expressing concern as a result, e.g., PL06: *"I would lose it, or I'd forget it because I would forget that I plugged it in a desktop computer because I can't see it."* Some saw this as a potential security threat, fearing that if the security key were always plugged in then an attacker could also use it.

Conversely, some participants equated a more substantial form factor with an increase in usability. Six participants commented on this aspect regarding the ePass, e.g., PL06: *"I would feel better about using it because it's like a USB."* Greater size, however, fuelled concerns about breaking the device, as it protrudes from the USB port, e.g., PL11: *"I did feel kind of like it could snap."*

The security keys all rely on a single touch-button interaction. However, this simple format created challenges for participants; all but one participant (PL10) failed to realize that the gold area on the YubiKey was the 'button' or 'gold disk' referred to by the Dropbox and Gmail user interface. In comparison, only a few participants failed to notice the 'button' on the ePass, e.g., PL02: *"because I just didn't see it as a button, it's flat."*

The form factor of the security keys also informed perceptions of when they could be used. Participants generally saw the ePass device as well suited to be carried around, whereas the SecureClick and the YubiKey devices were judged to be better suited to be attached to one computer.

#### 3.3.3 Rationalizing the Security Benefits of 2FA

Seven participants perceived that security keys provide additional security, but experienced challenges to articulate exactly how they provided added protection, and the threat they protected against, e.g., PL07: *"[it's] like an added protection, basically it is trying to identify it is you that is opening that account."* Only two participants recognized that security keys primarily defend against phishing attacks. One participant (PL03) perceived that having no visible association between devices and Web services adds further security, as opposed to for example bank tokens that are branded and thus more vulnerable to attacks: *"The issue [with bank tokens] is that it's all branded and everything so if it gets stolen, someone who's really desperate to have it work for him can actually do that, and for what I've seen with this, yes it's kind of branded but I could easily fit this in to my key holder and only I know what it's for."*

A few participants argued against the security provisions of security keys, for instance conveying that SMS-based OTPs were just as secure, but furthermore that Web service providers (such as Gmail and Yahoo) send real-time email notifications of any suspicious activity on the user account, e.g., PL07: *"[Gmail and Yahoo] send me a code and I have to log in and then they also send me 'You logged in from another device', so I guess because that happens automatically, I don't really have to bother myself, and then I feel a lot more secure when it happens."* One participant (PL04) added that locking security keys with biometric authentication would make it comparable to modern smartphones, for instance *"in case it gets lost [...] you kind of have that biometric control and power."*

## 4 Diary Study

To examine the fit of security keys with users' everyday practices, we conducted a diary study in January-February 2019. We focused specifically on the SecureClick security key since we were more knowledgeable about this device than the others, and could better support participants during day-to-day use (also discussed in Section 4.1.2). By asking participants to link a security key with personal online accounts, we hoped to capture more realistic usage data than a lab study could provide [31]. We chose a study time period of one week in order to minimize the burden on participants and hence an adverse effect on participation [33, 39].

### 4.1 Method

#### 4.1.1 Procedure

We recruited fifteen participants for the diary study, via a flyer/advert and electronic newsletters distributed across the sponsor university campus, and flyers shared with a nearby partner university. We also advertised the study on Twitter. There was no overlap in recruited participants with those recruited for the lab-based study.

Potential participants were directed to a pre-screen questionnaire, to provide basic details about the online services that they normally use. It was imperative to recruit participants that actively use U2F-compatible services (e.g., Dropbox, Facebook, GitHub, Google, Twitter, etc.). Participants should also actively use a desktop Web browser that supports

the use of security keys. No experience with security keys was necessary. Participants were compensated with their choice of a complimentary Amazon or Love2shop voucher worth £30 upon completing the study.

Each participant took part in a briefing session, lasting approximately 25 minutes. The experimenter followed a script to explain the study and would provide an information sheet and consent form; the participant was given time to read the documents before providing consent. To begin the study, we briefly discussed a participant's current authentication practices (revisiting the pre-screen responses). A researcher then guided the participant to set up a unique SecureClick security key (pre-paired with its Bluetooth Bridge) with up to two of their existing (U2F-compliant) accounts. Participants were free to set it up with their other accounts during their participation in the study if they wished.

Instructions were given on how to complete the diary journal (shown in the Appendix, Figures 6-7), and submit daily entries to the research team towards building rapport and to ensure participants remained motivated to complete the diary exercise [26, 39]. Participants who had not submitted their diary entries at the end of a day were reminded to do so the following day, in a single short message from the researcher. We managed communications with participants via email or WhatsApp, at the preference of each participant. After the diary exercise, each participant took part in a debrief interview, lasting approximately 15 minutes, to discuss their experiences and clarify uncertain entries provided during the diary exercise. Finally, online accounts linked to the SecureClick security key were restored to their initial state if requested by the participant. In addition to the financial incentive offered to take part in the study, participants were also offered to keep the SecureClick only at the time of leaving; 12 opted to do so.

Participants' answers during the brief and debrief interviews were audio-recorded to facilitate transcription at a later stage.

We finalized our study design by running a pilot study with one extra participant before the main study. We concluded that participants should use a personal computer in the briefing session as opposed to an unfamiliar device since we discovered this might serve as a deterrent to participation.

#### 4.1.2 Research Ethics

The diary study received ethical approval as part of the same project that included the lab-based study (Section 3).

Participants registered interest in taking part in the study using a pre-screen online form, managed from a survey platform operated from within the host university. We required a contact email address for the duration of the study and collected demographic information: age, gender, education level.

During the briefing session, up to two of a participant's online accounts were set up with a security key. Only researchers involved in the study had access to the dedicated email address and phone number, and we stored transcript data using a pseudonymous participant number.

To mitigate harm to study participants [19], we provided instructions on how to contact the research team with any issues during the study. The contact phone number for submitting diaries by WhatsApp, and for contacting researchers with issues, was terminated after the study.

#### 4.1.3 Demographics

We recruited fifteen participants for the study (5 female, 10 male). Participants ages were between 21- and 44-year old (median 24). Five were postgraduate, five were graduate, and five were up to the undergraduate level. Thirteen participants were from the host university and the remaining two from a partner university. Nine already had experience of using 2FA (either SMS-based OTPs or mobile-based authentication app). None of the participants had any familiarity with security keys.

#### 4.1.4 Limitations

The diary entries are self-reported data, which can be prone to under-reporting [34]. Participants may have adjusted their login behavior or diary completion to align with the perceived interests of the researchers. However, we had no leading hypothesis that could be leaked to the participant implicitly or otherwise that could prime participant behavior in one direction or another. Efforts were made to minimize interruption to participants, by asking for short entries in a structured table for each relevant event during the day, complemented with open-ended reflection only at the end of each day and at the end of the week (see Appendix). Our sample size is modest; however, our study design was structured to generate a useful volume of data irrespective of that. Crucially, none of our participants were familiar with security keys.

### 4.2 Results

We present the analysis of the diaries themselves and debrief interviews in the sections that follow. We refer to notable quotes from specific participants using numeric identifiers: e.g., PD##.

#### 4.2.1 Diary Entry Analysis

**Overall Activity**: We recorded 643 diary entries across all participants. The median number of diary entries per participant was 38.

The median number of Web services that a participant registered their security key with was 3 ($IQR = 2$). As illustrated in Figure 3, the services for which participants reported the most frequent logins were: Gmail (321 events, 50%), Facebook (114, 18%), Dropbox (95, 15%) and

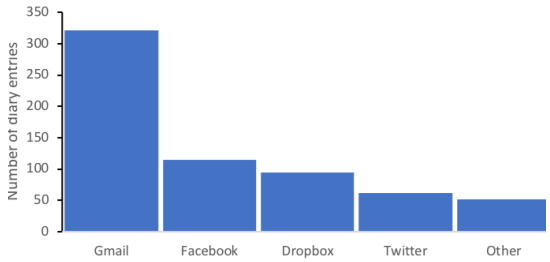



Figure 3: Chart of the services for which participants chose to register their security key and for which we recorded login events.

Twitter (61, 10%).

**Locations of Events:** The median number of different locations that participants used the security key was 4 ($IQR = 4$). Participants reported that the vast majority of login events took place in a home environment (396 events, 62%); followed by the workplace (80, 12%); whilst in transit (e.g., on trains, buses) (52, 8%) and at a university (57, 8%); in public spaces, e.g., cafes (48, 7%); and finally, at a friends' house (10, 2%).

**Computer Use:** The median number of computers that participants accessed accounts from where the security key was enabled was 2 ($IQR = 2$). The most common device used with the security key was a personal laptop (345 events, 54%), followed by an own mobile device (191, 30%), a personal tablet (41, 6%), a work desktop (32, 5%), and public computers (13, 2%). Other devices comprising less than 1% of accesses included the devices of friends/family.

**Device Management and Portability:** Concerning how participants managed the two parts of the SecureClick, eight participants always kept both the button and USB parts of the SecureClick together, whereas the remaining seven kept them separately. There was a mix of attitudes regarding where participants kept the parts: keyring (button part only), laptop (USB part only), wallet, original box, clear plastic case, bag, pocket, safe place at home or work.

In terms of portability, ten participants generally carried both parts of the SecureClick and thus always had access to it if needed. On the other hand, three participants carried only the button part and left the USB part plugged in their laptop at home at all times, and the remaining two participants always left both parts of the SecureClick in a safe place at home.

**Login Methods:** Table 3 illustrates the most common means by which participants accessed security key-enabled accounts during the study. The most common type of login event recorded was where users utilized the 'automatic login' functionality of a Web service (63%). This is where a Web service remembers a successful login on a specific device and does not prompt the user to authenticate again for a time, such as 30 days. Thus the user is logged in transparently and without any authentication friction. The combination of a security key and password appeared in only 28% of the captured logins. A password and alternative 2FA, such as SMS, was used 5% of the time, and circumstances were possible where users reported being able to access a service with only a password – a possibility that appears specific to Gmail. Participants abandoned 2% of the sessions due to issues with accessing a service.

| **Login Type** | **Percent** |
|---|---|
| Automatic login | 63% |
| Password & Security Key | 28% |
| Password & Other 2FA | 5% |
| Password only | 2% |
| Abandoned sessions | 2% |

Table 3: Frequency of the different types of captured logins.

Figure 4 shows the 2FA login methods used on each day of the study across all participants as a proportion of all login events, for online accounts with an enabled security key. Usage of automatic login increased over time at the expense of security keys – automatic login and the combination of password and security key were used 38% and 44% of the time respectively on the first day of the study, whereas the figures were 75% and 16% respectively at the end of the week.

**Satisfaction:** At the end of each day participants were asked to respond to the question *"On a scale of 1 (very bad) to 9 (very good), how would you rate your experience of using the security key today?"* The median response was 7 ($IQR = 3$), and there was no discernible relationship with these scores and the progression of the study. An example of a free-text response to a day with a score of 3 would be (PD07) *"annoyances started when attempting to log in using a mobile phone, Git pushing from Ubuntu terminal, as well as logging in with different browsers on public computers like Firefox."* The example of GitHub is of specific interest since a security key can be enabled on GitHub for login as long as SMS-based OTPs are also enabled. Then, when using the command line interface, if 2FA is enabled, the user must register an SSH key to authenticate repository updates (since the browser cannot capture 2FA from the command line). Another low score (4) included the comment: *"[Facebook] still sent an SMS message. This doesn't feel particularly secure if every time I use the security key, I get an SMS. Why not just use the phone if it's going to communicate with the phone, anyway?"* (PD01). The comment refers to Facebook's practice of sending an SMS-based OTP even if a user is being prompted to use a security key. An example of comments associated with a high score (9 out of 9) was (PD03) *"I only realized now this solution is excellent when using public computers."*

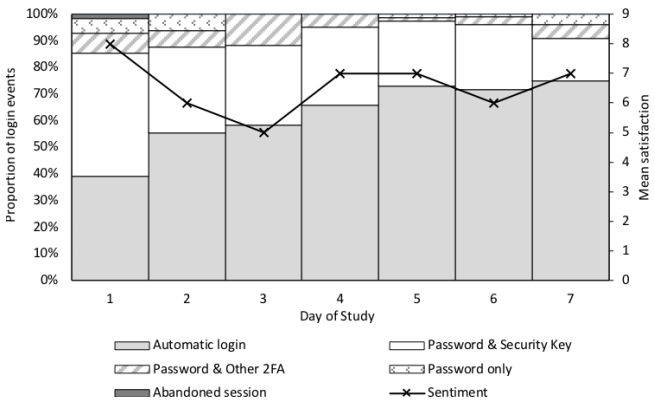



Figure 4: Proportion of login methods used on each day of the study across all participants, with Sentiment, Very Bad (1) to Very Good (9), on right-hand y-axis. Usage of 'automatic login' increases over time at the expense of security keys. Sentiment towards the security key stayed relatively constant over the course of the week.

The sentiment towards the security key stayed relatively constant over the week, as highlighted in Figure 4. We noted this effect despite the usage of the security key decreasing over time.

### 4.2.2 Qualitative Results

We analyzed the qualitative data using thematic analysis [12]. All participant responses were coded in several stages by one researcher, initially as low-level labels, moving to higher-level analytical categories. We identified five high-level categories – device form factor was omitted as a repeat of lab study findings, leaving the four themes presented in the following sub-sections – and 16 sublevel codes.

**Threats, Context, and the Purpose of FIDO U2F**

In terms of perceived threats, participants were generally not concerned about losing their passwords via phishing emails, with only one participant (PD05) in contrast conveying that they were *"massively worried"* about it. Five participants were worried about using public machines (PD02, PD03, PD04, PD05, PD13), attributing this to potential loss of credentials via malware or shoulder surfing. Three participants (PD05, PD09, PD15) mentioned concerns about losing their credentials when using public WiFi. Two participants (PD05, PD10) were worried about losing their laptop or having it stolen, lest it permits an attacker to access their online accounts.

In terms of general authentication practices away from the study, participants predominantly used 1FA to access their online accounts, with 2FA via SMS-based OTPs seldom used when forced to do so by specific Web services, e.g., banks or work Virtual Private Network (VPN). To facilitate access to Web services, eleven participants reported using a password manager (dedicated, or credentials saved in the browser), with a further three using automatic login. The remaining two participants reported re-entering their credentials each time they accessed their online accounts.

Three participants mentioned proactively taking extra steps to secure their online accounts. PD01 implemented a bespoke password manager, as they did not trust commercially available tools. PD02 reported using a widely available password manager to increase the *"entropy"* of their passwords, also only using their own 'trusted' personal machine, having *"hardened it and [I] don't let people put random USB in."*

Thirteen participants perceived the security key as useful only in specific contexts, generally to protect accounts holding sensitive information (e.g., emails, work, banking). The remaining two participants did not see a use for the security key. Five participants thought it could be useful when accessing accounts using public machines or another person's machine, although there was a concern that the machine owners would conversely be uneasy about allowing an 'unknown' technology to interact with their machine. Two participants (PD06, PD14) speculated that the security key could be useful to secure work-related accounts on a specific device. PD03 also mentioned that security keys could be useful to secure physical objects: *"This device, I don't know if you could use it in a way to secure storage."*

**Security Key as a Perceived Barrier to Login**

Three participants (PD01, PD08, PD12) explained that the security key affected or reduced their inclination to access online services because *"it does make a conscious barrier between you [when you] log in onto a site, because it's a different action."* (PD01). This effect was particularly noted during access of social media sites: *"It has reduced my usage by quite a lot [...] I think I had at the back of my mind that I would need to go back in my bag, get the key out, put it in, go onto the thing."* (PD01); for some this barrier was not unwelcome, e.g., PD08: *"I wasted less time on Facebook whenever I was in the library."* Four other participants (PD05, PD09, PD10, PD11) thought that the login delay introduced by using the security key could be frustrating. Issues with 'authentication fatigue' have been reported elsewhere as factors in employees reducing their use of computers [40]. Only one participant (PD05) reported a perceived increase in their usage of Google services as a result of using the security key: *"I've already started to save some stuff to Google, which I wasn't doing before, because it felt safer now."*

**Challenges in Configuring and Using FIDO U2F**

An inability to make the security key work with a mobile phone was a recurring issue affecting six of the fifteen participants.

Other set up issues concerned poor or complete lack of

phone reception when setting up mandatory backup authentication mechanisms (PD02, PD03); a participant's browser not supporting U2F by default (PD05, PD11); an inability to find an option to set up the security key with Google (PD01, PD13) or specific applications (Thunderbird (PD04), Apple Mail (PD12)); and the participant's computer OS not supporting U2F by default (PD02).

Usage issues specific to the SecureClick were sporadic. Two participants (PD06, PD12) experienced inability to, or had issues with, getting the SecureClick to work on a new computer due to problems 'installing' the USB part when plugging it into the new computer. Two other participants (PD09, PD15) complained about being unable to leave the USB part plugged into their computer, as they have to pull the dongle out and put it back in for it to work again after restarting their computer. Two participants (PD11, PD12), where the SecureClick ran very low on battery and effectively stopped working, failed to recognize this was the case, as the light feedback still operated as usual in these instances (where insufficient battery life for a security key has been seen as an issue elsewhere [40]).

Other general challenges to using security keys lay in the limited support of U2F amongst browsers (PD05, PD07, PD10, PD15) and operating systems (PD07). Two participants mentioned the barrier to the use of security keys with Git CLI for GitHub (PD02, PD07).

Some of these issues caused two participants to remove the SecureClick from some of their accounts, e.g., PD07: *"I disabled it on GitHub, because the pushing and pulling part was just getting a bit annoying."*

**User Choices Lack Support**

User selections of authentication mechanisms were often not respected or persisted. Participants were initially frustrated if provided 2FA options did not meet their expectation of needing to use the security key. Eight participants were unaware that some Web services (e.g., Gmail) enabled a 'remember this security key' option by default on the first login, sometimes leaving participants puzzled as to why they were not prompted for the security key again during subsequent login events, e.g., PD10: *"I think it just surprised me when it did it, and I'm not 100% sure if it was done by my computer or if it was the token that initially did that."* Two participants (PD01, PD04) had issues with some Web services sending them an SMS-based OTP at the same time that they were using the security key, e.g., PD01: *"Facebook spammed my iPhone with texts suggesting I needed a code - even when the key was working."*.

Six participants thought that services offering a 'remember this security key' option, or sending an SMS-based OTP at the same time as a security key was being used, rendering the keys redundant, e.g., PD04: *"Facebook always sends an SMS code at the same time, so I didn't need the key to do it, so the key feels useless at that point."*

Three participants (PD02, PD06, PD09) mentioned that using a security key should come with weaker or less stringent password policies. Regarding the choice of login methods, four participants (PD02, PD07, PD09, PD13) preferred persistent login sessions, two (PD05, PD08) did not, and three (PD06, PD10, PD12) would make different decisions depending on the context. Two participants (PD07, PD12) also mentioned that they would like a choice of different login policies for different devices. These results allude to the decisions about whether to use a particular combination of 2FA technologies being more complicated than which two to use, but how to combine them to reach a level of security that is satisfactory to the user.

# 5 Discussion

## 5.1 Comparisons between 2FA techniques

Revisiting our overarching research question (Section 2.3), our lab-based study evaluated a diverse cross-vendor set of security keys alongside SMS-based OTPs. The security key setup process is generally not efficient for novice users to complete [38], but we found that the devices were deceptively heterogeneous, and created their own specific challenges. Setup times were considerably larger than login times for the security keys. The median overall setup time was 146 seconds ($IQR = 101$), and the median login time was 30 seconds ($IQR = 17$). No particular device was significantly more successful than another in enabling greater efficiency, despite device-specific features being known to impact efficiency for users, e.g., the Bluetooth pairing required by the SecureClick, and the (small) size of the YubiKey. The ePass was free of both issues, but required drivers to be installed (on Windows platforms) before use (creating and contributing to setup delays). These findings challenge the often remarked claims that these are simple, *'one tap'* devices.

SMS-based OTPs were never rated below acceptable by participants through SUS ratings, whereas security keys often were. The high ratings that participants provided for SMS-based OTPs were not in line with measurements of time efficiency during our laboratory study. This result could be symptomatic of participants trusting the familiar SMS technology, or could indicate that users anticipate security keys impacting on convenience and account recovery. Also, configuring backup 2FA (typically SMS-based OTPs) was often a pre-requisite for setting up security keys, which could have led participants to attach more significance to the role of SMS, rather than security keys.

## 5.2 Everyday Experiences of Security Keys

Prior work [38] has reported that users were generally satisfied when using a security key; we obtained similar results through SUS ratings: mean=75.83 ($SD = 14.81$, 95% $CI = \pm 7.49$), or

'acceptable' [9]. However, participants only used the security keys in 28% of the recorded login events in which security keys were active. On the final day of the diary study, the daily usage of the security key had declined as a proportion of all login events by 50% compared to the first day. This decrease could partly explain why prior work highlights acceptable user satisfaction with security keys in field studies [38], yet greater challenges in lab-based studies focusing on the keys themselves [16].

Participants were generally using or willing to use 2FA with Web services, at least for accounts that they deemed to be critical. However, there was a perceived risk of being locked out of one's account, should the second factor be unavailable when needed (the activity of locating the key, as in Figure 5 – see Appendix). This risk was a major concern for participants, and the form factor of security keys may exacerbate such fears. Participants perceived some security keys as more suited for use in one place only (e.g., at home, or in a work environment), whereas others were judged acceptable to be carried around and used for login from different computers on the move. Some participants felt that it was inevitable to require ownership of several security keys for this reason. However, the diversity of such suggestions hints that users struggled to spot an identifiable 'universal' usage proposition for the security keys. It may be that a use case for security keys is as devices for infrequent use in bootstrapping a set of trusted devices, for subsequent transparent logins. If distinct use cases were to emerge, this would require device manufacturers to set different expectations about how users should optimally use the device.

### 5.3 Service Providers and Managing Friction

Security keys are user-centric [10] technologies that are decentralized. As such, there is no natural recovery mechanism that can be provided by the service provider should a device be lost, except to provide a toolbox of ready 2FA alternatives. Service providers encounter a conflict between reducing authentication friction for their customers to access services easily, and to enable users to protect their accounts. The same conflict has been noted with alphanumeric password strength for online services, where those with the largest customer bases had the least stringent password requirements [22].

But there may be risks to completely removing friction from security key usage. Specifically, user trust in the devices may decline due to the way that the user interface prioritizes other 2FA options. As an example, if 2FA is enabled, Facebook sends an SMS to the user at the point of login, even if the user previously selected to use a security key. Similarly, Gmail occasionally requests only a password to login to a device where the user used a security key in the past and, indeed, security keys are at the bottom of the list of alternative 2FA methods in the choice architecture for Gmail 2FA. Finally, '*Remember me'* was a feature that constituted the majority of service accesses in our diary study. Each of these examples illustrate how the user perception of the importance of pressing the button on a security key is undermined since that preference is under constant challenge by the presentation of 2FA alternatives at crucial moments. These events may act like a 'nudge' of the user towards a preferred 2FA [37], rather than the display of a choice architecture that promotes informed decisions for a particular user [13].

The transition towards FIDO2 [4] may alleviate some of these challenges, through closer integration of U2F with mobile devices. In the long-term, it may be that these standards are necessary not only to move toward seamless mobile device support, but also to support service providers to optimize the design of their infrastructure around future iterations of U2F or even decentralized identifiers from emerging decentralized identity schemes [20].

## 6 Conclusion

Security keys are 2FA technologies that are resistant to phishing, whereas ubiquitous SMS codes are not. However, uptake of security keys for general Web browsing is generally low. In this paper, we conducted two empirical studies to better understand the user experience of security keys for purposes of everyday Web authentication.

Firstly, in a laboratory study, we evaluated a diverse cross-vendor set of security keys alongside SMS-based OTPs, to capture factors affecting the usability and security perceptions of security keys during setup and login. We found that the setup time for security keys was considerably greater than login time. Also, SMS-based OTPs were never rated below 'acceptable' by participants using an SUS scale, whereas security keys often were.

Secondly, we conducted a diary study over one week, to capture user experience challenges encountered in everyday use of a security key. We found that only 28% of accesses to security key-enabled online accounts involved pressing a button on a security key, and use of a security key decreased as a proportion of all account accesses as the study progressed. Inadvertently nudging users away from explicit use of security keys likely erodes the perception of utility of security keys which is seen in prior work [16].

Our research demonstrates the importance of considering security key usage in the broader context of other competing 2FA technologies and the nature of 2FA choice architectures provided by Web services.

## Acknowledgments

We would like to thank the SOUPS reviewers for their comments and support in preparing the paper for the conference. Stéphane Ciolino was supported in part by OneSpan. Study incentive and security key costs were supported by OneSpan.

## References


[1] Seb Aebischer, Claudio Dettoni, Graeme Jenkinson, Kat Krol, David Llewellyn-Jones, Toshiyuki Masui, and Frank Stajano. Pico in the Wild: Replacing Passwords, One Site at a Time. In *Proc. European Workshop on Usable Security (EuroUSEC 2017)*. Internet Society, 2017. URL: https://www.internetsociety.org/sites/default/files/eurousec2017_17_Aebischer_paper.pdf, doi:10.14722/eurousec.2017.23017.

[2] FIDO Alliance. About The FIDO Alliance. URL: https://fidoalliance.org/about/overview/.

[3] FIDO Alliance. Approach & Vision. URL: https://fidoalliance.org/approach-vision/.

[4] FIDO Alliance. FIDO2: Moving the World Beyond Passwords using WebAuthn & CTAP. URL: https://fidoalliance.org/fido2/.

[5] FIDO Alliance. 2017 State of Authentication Report, October 2017. URL: https://fidoalliance.org/2017-state-authentication-report/.

[6] FIDO Alliance. MakeUseOf: It's Time to Stop Using SMS and 2FA Apps for Two-Factor Authentication, January 2018. URL: https://fidoalliance.org/time-stop-using-sms-2fa-apps-two-factor-authentication/.

[7] M. M. Althobaiti and P. Mayhew. Security and usability of authenticating process of online banking: User experience study. In *2014 International Carnahan Conference on Security Technology (ICCST)*, pages 1–6, October 2014. doi:10.1109/CCST.2014.6986978.

[8] All Things Auth. SMS: The most popular and least secure 2FA method, February 2018. URL: https://www.allthingsauth.com/2018/02/27/sms-the-most-popular-and-least-secure-2fa-method/.

[9] Aaron Bangor. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *Journal of Usability Studies*, 4(3):114–123, May 2009.

[10] Abhilasha Bhargav-Spantzely, Jan Camenisch, Thomas Gross, and Dieter Sommer. User centricity: a taxonomy and open issues. In *Proceedings of the second ACM workshop on Digital identity management - DIM '06*, page 1, New York, New York, USA, 2006. ACM Press. URL: http://portal.acm.org/citation.cfm?doid=1179529.1179531, doi:10.1145/1179529.1179531.

[11] Joseph Bonneau, Cormac Herley, Paul C. van Oorschot, and Frank Stajano. The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes. In *IEEE Symp. on Security and Privacy*, pages 553–567, May 2012. URL: http://ieeexplore.ieee.org/document/6234436/, doi:10.1109/SP.2012.44.

[12] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, January 2006. URL: https://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa, doi:10.1191/1478088706qp063oa.

[13] Pamela Briggs, Debbie Jeske, and Lynne Coventry. Behavior change interventions for cybersecurity. In *Behavior Change Research and Theory*, pages 115–136. Elsevier, 2017.

[14] Elizabeth Charters. The Use of Think-aloud Methods in Qualitative Research An Introduction to Think-aloud Methods. *Brock Education Journal*, 12(2), July 2003. URL: https://brock.scholarsportal.info/journals/brocked/home/article/view/38, doi:10.26522/brocked.v12i2.38.

[15] Jessica Colnago, Summer Devlin, Maggie Oates, Chelse Swoopes, Lujo Bauer, Lorrie Cranor, and Nicolas Christin. "It's not actually that horrible": Exploring Adoption of Two-Factor Authentication at a University. In *CHI 2018*, pages 1–11, Montreal, QC, Canada, April 2018. ACM Press. URL: http://dl.acm.org/citation.cfm?doid=3173574.3174030, doi:10.1145/3173574.3174030.

[16] Sanchari Das, Andrew Dingman, and L Jean Camp. Why Johnny Doesn't Use Two Factor A Two-Phase Usability Study of the FIDO U2F Security Key. *Preproceedings Financial Cryptography and Data Security 2018*, 2018.

[17] Sanchari Das, Gianpaolo Russo, Andrew Dingman, Jayati Dev, Olivia Kenny, and L Jean Camp. A Qualitative Study on Usability and Acceptability of Yubico Security Key. *Proceedings of , Florida, USA, December (STAST 2017)*, December 2017.

[18] Emiliano De Cristofaro, Honglu Du, Julien Freudiger, and Greg Norcie. A Comparative Usability Study of Two-Factor Authentication. In *Proceedings 2014 Workshop on Usable Security*, San Diego, CA, 2014. Internet Society. URL: https://www.ndss-symposium.org/ndss2014/workshop-usable-security-usec-2014-programme/comparative-usability-study-two-factor-authentication, doi:10.14722/usec.2014.23025.

[19] David Dittrich, Erin Kenneally, et al. The Menlo Report: Ethical principles guiding information and communication technology research. Technical report, US Department of Homeland Security, 2012.

[20] P. Dunphy and F. A. P. Petitcolas. A First Look at Identity Management Schemes on the Blockchain. *IEEE Security Privacy*, 16(4):20–29, July 2018. doi:10.1109/MSP.2018.3111247.

[21] Duo. Bringing U2F to the Masses. URL: https://duo.com/blog/bringing-u2f-to-the-masses.

[22] Dinei Florêncio and Cormac Herley. Where Do Security Policies Come From? In *Proceedings of the Sixth Symposium on Usable Privacy and Security*, SOUPS '10, pages 10:1–10:14, New York, NY, USA, 2010. ACM. URL: http://doi.acm.org/10.1145/1837110.1837124, doi:10.1145/1837110.1837124.

[23] Bennett Garner. Why 2FA Matters & the Best Types of 2FA, April 2018. URL: https://coincentral.com/why-2fa-matters-the-best-types-of-2fa/.

[24] Greg Guest, Arwen Bunce, and Laura Johnson. How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. *Field Methods*, 18(1):59–82, February 2006. URL: http://journals.sagepub.com/doi/10.1177/1525822X05279903, doi:10.1177/1525822X05279903.

[25] Nancie Gunson, Diarmid Marshall, Hazel Morton, and Mervyn Jack. User perceptions of security and usability of single-factor and two-factor authentication in automated telephone banking. *Computers & Security*, 30(4):208–220, June 2011. URL: http://www.sciencedirect.com/science/article/pii/S0167404810001148, doi:10.1016/j.cose.2010.12.001.

[26] P. G. Inglesant and M. A. Sasse. Studying Password Use in the Wild: Practical Problems and Possible Solutions. In *Proceedings of the Sixth Symposium on Usable Privacy and Security*, 2010. URL: https://cups.cs.cmu.edu/soups/2010/user_papers/Inglesant_passwords_in_wild_USER2010.pdf.

[27] Kaspersky. SMS-based two-factor authentication is not safe - consider these alternative 2FA methods instead, October 2018. URL: https://www.kaspersky.co.uk/blog/2fa-practical-guide/14589/.

[28] Brian Krebs. Google: Security keys neutralized employee phishing, 2018. URL: https://krebsonsecurity.com/2018/07/google-security-keys-neutralized-employee-phishing/.

[29] Kat Krol, Simon Parkin, and M. Angela Sasse. Better the Devil You Know: A User Study of Two CAPTCHAs and a Possible Replacement Technology. In *Proceedings of NDSS Workshop on Usable Security (USEC 2016)*, San Diego, CA, USA, 2016. Internet Society. URL: https://wp.internetsociety.org/ndss/wp-content/uploads/sites/25/2017/09/better-the-devil-you-know-user-study-of-two-captchas-a-possible-replacement-technology.pdf, doi:10.14722/usec.2016.23013.

[30] Kat Krol, Eleni Philippou, Emiliano De Cristofaro, and M. Angela Sasse. "They brought in the horrible key ring thing!" Analysing the Usability of Two-Factor Authentication in UK Online Banking. In *USEC '15*, San Diego, CA, USA, February 2015. Internet Society. URL: https://www.ndss-symposium.org/ndss2015/ndss-2015-usec-programme/they-brought-horrible-key-ring-thing-analysing-usability-two-factor-authentication-uk-online, doi:10.14722/usec.2015.23001.

[31] Kat Krol, Jonathan M Spring, Simon Parkin, and M Angela Sasse. Towards robust experimental design for user studies in security and privacy. In *The LASER Workshop: Learning from Authoritative Security Experiment Results (LASER 2016)*, pages 21–31, San Jose, CA, May 2016. USENIX Association.

[32] Juan Lang, Alexei Czeskis, Dirk Balfanz, Marius Schilder, and Sampath Srinivas. Security Keys: Practical Cryptographic Second Factors for the Modern Web. In *Financial Cryptography and Data Security*, Lecture Notes in Computer Science, pages 422–440. Springer, Berlin, Heidelberg, February 2016. URL: https://link.springer.com/chapter/10.1007/978-3-662-54970-4_25, doi:10.1007/978-3-662-54970-4_25.

[33] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. *Research methods in human-computer interaction*. Elsevier, Cambridge, MA, 2nd edition edition, 2017. OCLC: 1030364616.

[34] Shrirang Mare, Mary Baker, and Jeremy Gummeson. A Study of Authentication in Daily Life. In *Proceedings of the Twelfth USENIX Conference on Usable Privacy and Security*, SOUPS'16, pages 189–206, Berkeley, CA, USA, 2016. USENIX Association. URL: http://dl.acm.org/citation.cfm?id=3235895.3235912.

[35] National Cyber Security Centre (NCSC). Setting up two-factor authentication (2FA), 2018. URL: https://www.ncsc.gov.uk/guidance/setting-two-factor-authentication-2fa.

[36] Jakob Nielsen. Finding usability problems through heuristic evaluation. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 373–380. ACM, 1992.

[37] Karen Renaud and Verena Zimmermann. Ethical guidelines for nudging in information security & privacy. *International Journal of Human-Computer Studies*, 120:22–35, 2018.

[38] Joshua Reynolds, Trevor Smith, Ken Reese, Luke Dickinson, Scott Ruoti, and Kent Seamons. A Tale of Two Studies: The Best and Worst of YubiKey Usability. In *2018 IEEE Symposium on Security and Privacy (SP)*, pages 872–888, San Francisco, CA, May 2018. IEEE. URL: https://ieeexplore.ieee.org/document/8418643/, doi:10.1109/SP.2018.00067.

[39] John Rieman. The Diary Study: A Workplace-oriented Research Tool to Guide Laboratory Efforts. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*, CHI '93, pages 321–326, New York, NY, USA, 1993. ACM. URL: http://doi.acm.org/10.1145/169059.169255, doi:10.1145/169059.169255.

[40] M. Angela Sasse, Michelle Steves, Kat Krol, and Dana Chisnell. The Great Authentication Fatigue - And How to Overcome It. In P. L. Patrick Rau, editor, *Cross-Cultural Design*, Lecture Notes in Computer Science, pages 228–239. Springer International Publishing, 2014.

[41] Jonathan M. Spring and Phyllis Illari. Building General Knowledge of Mechanisms in Information Security. *Philosophy & Technology*, Sep 2018. URL: https://doi.org/10.1007/s13347-018-0329-z, doi:10.1007/s13347-018-0329-z.

[42] Michelle Steves, Dana Chisnell, Angela Sasse, Kat Krol, Mary Theofanos, and Hannah Wald. Report: Authentication Diary Study. Technical Report NIST IR 7983, National Institute of Standards and Technology, February 2014. URL: https://nvlpubs.nist.gov/nistpubs/ir/2014/NIST.IR.7983.pdf, doi:10.6028/NIST.IR.7983.

[43] Jake Weidman and Jens Grossklags. I Like It, but I Hate It: Employee Perceptions Towards an Institutional Transition to BYOD Second-Factor Authentication. In *Proceedings of the 33rd Annual Computer Security Applications Conference on - ACSAC 2017*, pages 212–224, Orlando, FL, USA, 2017. ACM Press. URL: http://dl.acm.org/citation.cfm?doid=3134600.3134629, doi:10.1145/3134600.3134629.

[44] Catherine S. Weir, Gary Douglas, Martin Carruthers, and Mervyn Jack. User perceptions of security, convenience and usability for ebanking authentication tokens. *Computers & Security*, 28(1-2):47–62, February 2009. URL: http://linkinghub.elsevier.com/retrieve/pii/S0167404808000941, doi:10.1016/j.cose.2008.09.008.

[45] Catherine S. Weir, Gary Douglas, Tim Richardson, and Mervyn Jack. Usable security: User preferences for authentication methods in eBanking and the effects of experience. *Interacting with Computers*, 22(3):153–164, May 2010. URL: https://academic.oup.com/iwc/article-lookup/doi/10.1016/j.intcom.2009.10.001, doi:10.1016/j.intcom.2009.10.001.

[46] Wired. How to Protect Yourself Against a SIM Swap Attack. URL: https://www.wired.com/story/sim-swap-attack-defend-phone/.

[47] Yubico. Works with YubiKey. URL: https://www.yubico.com/solutions/.

[48] Yubico. OTP vs. U2F: Strong To Stronger, February 2016. URL: https://www.yubico.com/2016/02/otp-vs-u2f-strong-to-stronger/.

## Appendices

## A Lab-Study Task: 2FA Using *[Tested Mechanism]* on Laptop

- ['Set-up phase'] On laptop, ask participants to:
  - [SecureClick only] Install OneSpan DIGIPASS SecureClick Manager and pair SecureClick with its Bluetooth Bridge.
  - open Chrome.
  - login onto Web service.
  - set up 2FA using *[Tested Mechanism]* on Web service.
  - logout of Web service.
  - close Chrome.
- ['Login' phase] On laptop, ask participants to:
  - open Chrome.
  - login onto Web service.
  - logout of Web service.
  - close Chrome.
- [SUS] Ask participants to fill in the SUS questionnaire about their experience of 2FA using *[Tested Mechanism]* on laptop.

## B FIDO U2F Mechanisms

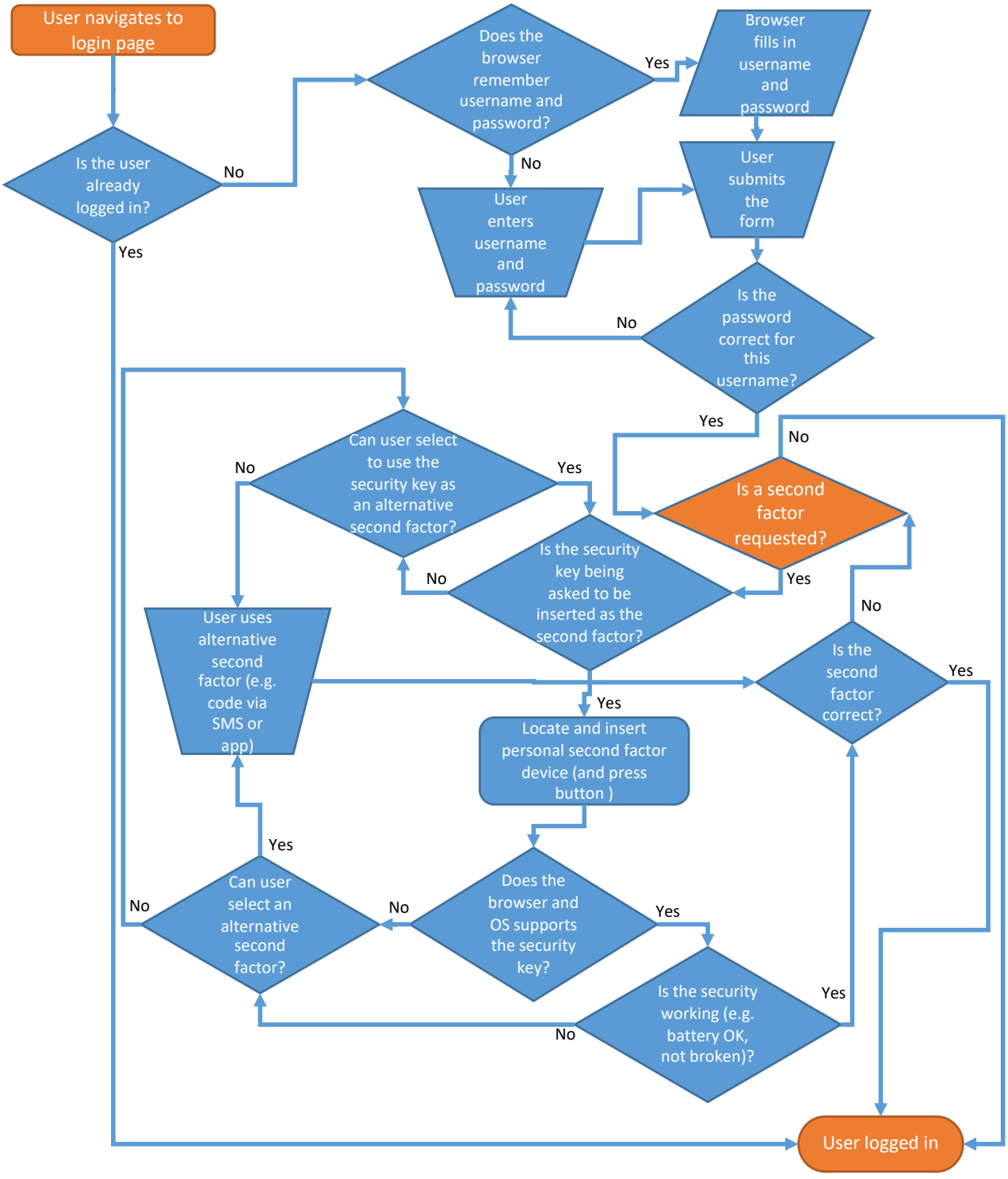



Figure 5: Mechanisms and activities in the FIDO U2F study.

# C Diary Forms

## Per Access

**For each attempted access (successful or otherwise) to a web service where the DIGIPASS SecureClick (security key) is enabled,** please fill in the following information in a new row.

| # | Time | Web service you were accessing *(Gmail, Facebook, Twitter, Dropbox, GitHub, etc)* | Location you were accessing it from *(home, work, internet café, etc)* | Device you were accessing it from *(personal laptop, work desktop, mobile, etc)* | Successful access? *(Yes/No)* | How many unsuccessful attempts did you have before successful access or abandoning? | If the access was successful, which authentication method(s) did you use to access your account? *(Tick all that apply)* | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Security key | Code (via SMS, email, other) | Password | Nothing. Automatic login |
| 0 | 13:15 | Twitter | Library | Public desktop | Yes | 0 | ☒ | ☐ | ☒ | ☐ |
| 1 | | | | | | | ☐ | ☐ | ☐ | ☐ |
| 2 | | | | | | | ☐ | ☐ | ☐ | ☐ |
| 3 | | | | | | | ☐ | ☐ | ☐ | ☐ |
| 4 | | | | | | | ☐ | ☐ | ☐ | ☐ |
| 5 | | | | | | | ☐ | ☐ | ☐ | ☐ |
| 6 | | | | | | | ☐ | ☐ | ☐ | ☐ |
| 7 | | | | | | | ☐ | ☐ | ☐ | ☐ |
| 8 | | | | | | | ☐ | ☐ | ☐ | ☐ |
| 9 | | | | | | | ☐ | ☐ | ☐ | ☐ |
| 10 | | | | | | | ☐ | ☐ | ☐ | ☐ |

*(Continues next page if needed)*

## At the End of the Day

**1. Which of the following, if any, did you experience today? *(Please tick any that apply and specify where required)***

a. I could readily get hold of the security key whenever I wanted to use it: ☐ Yes ☐ No

If you misplaced or could not readily get hold of some part(s) of the security key when you wanted to use it, please specify...

☐ Button part ☐ USB part For how many hours:

b. The security key always worked as I expected throughout the day: ☐ Yes ☐ No, specify:

c. I used an alternative authentication method to the security key during at least one access attempt: ☐ Yes ☐ No

If Yes, please specify which method(s) you used instead of using the security key:

☐ Code via SMS ☐ Code via mobile authenticator app ☐ Other, please specify:

d. I felt it was quick to get the security key ready to use, and to use it: ☐ Yes ☐ No How long did it typically took:

e. I found the feedback (e.g. light) from the security key clear: ☐ Yes ☐ No, please specify:

f. Other, please specify:

**2. What best describes where your security key has been today? *(Please tick any that apply and specify where required)***

a. Part(s) of the security key have been on my person: ☐ Yes ☐ No

If Yes, please specify: ☐ Button part ☐ USB part For how many hours:

b. Part(s) of the security key have been somewhere safe, but not on my person: ☐ Yes ☐ No

If Yes, please specify: ☐ Button part ☐ USB part For how many hours:

c. Other, please specify:

**3. Please leave any further comments you may have about your experience of using the security key today.**

**4. On a scale of 1 to 9, how would you rate your experience of using the security key today? *(Please tick which best applies)***

| Very bad | | | | Neither bad nor good | | | | Very good |
|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| | | | | | | | | |

Figure 6: Daily diary form for the FIDO U2F diary study.

**End of Week Questionnaire**

We are now focusing on your overall personal experience of using the security key during the last 7 days. Please keep this in mind when answering the following questions.

**1. Did you set up (or wanted to set up) the security key with any other web service(s)?** ☐ Yes ☐ No

If Yes, please specify with which web service(s):

**2. Did you remove (or wanted to remove) the security key from any web service(s)?** ☐ Yes ☐ No

If Yes, please specify with which web service(s):

**3. Did using the security key affect the way you access web services (e.g. accessing web services less/more often than usual, or from different locations and/or devices)?** ☐ Yes ☐ No

If Yes, please specify how it affected your behavior:

**4. Thinking specifically about web services where the security key is enabled, some offer an option to 'remember' the security key on devices you trust. Did you use this option?** ☐ Yes ☐ No ☐ Don't know

If Yes, please specify with which web service(s):

**5. Thinking specifically about web services where the security key is enabled, did you need to look at instructions or get help from someone to access your account(s)?** ☐ Yes ☐ No ☐ Don't know

**6. You have now been using the security key for a week. Overall, on a scale of 1 to 9, how would you rate your experience of using the security key? *(Please tick which best applies)***

| Very bad | | | | Neither bad nor good | | | | Very good |
|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| | | | | | | | | |

**7. Following this study, to what extent would you see yourself using a security key if you had one? *(Please tick which best applies)***

☐ I would never use it ☐ I would use it only in some specific contexts ☐ I would use it in any context

**8. Please leave any further comments you may have about your overall experience of using the security key.**

Figure 7: End-of-week diary form for the FIDO U2F diary study.

# A Usability Study of Five Two-Factor Authentication Methods

Ken Reese, Trevor Smith, Jonathan Dutson, Jonathan Armknecht, Jacob Cameron, Kent Seamons
*Brigham Young University*

## Abstract

Two-factor authentication (2FA) defends against account compromise. An account secured with 2FA typically requires an individual to authenticate using something they know—typically a password—as well as something they have, such as a cell phone or hardware token. Many 2FA methods in widespread use today have not been subjected to adequate usability testing. Furthermore, previous 2FA usability research is difficult to compare due to widely-varying contexts across different studies. We conducted a two-week, between-subjects usability study of five common 2FA methods with 72 participants, collecting both quantitative and qualitative data. Participants logged into a simulated banking website nearly every day using 2FA and completed an assigned task. Participants generally gave high marks to the methods studied, and many expressed an interest in using 2FA to provide more security for their sensitive online accounts. We also conducted a within-subjects laboratory study with 30 participants to assess the general usability of the setup procedure for the five methods. While a few participants experienced difficulty setting up a hardware token and a one-time password, in general, users found the methods easy to set up.

## 1 Introduction

Passwords are the most widespread form of user authentication on the web today [9]. Although many password-replacement schemes have been proposed, none of them compete with the deployability and usability of passwords [8]. Recently, large service providers, including Google, Facebook, and Microsoft, have deployed an optional 2FA layer as part of their authentication processes to defend against account compromise. Two-factor authentication requires users to present two of the following types of authentication factors:

1. Something they *know* (traditionally a password)
2. Something they *have* (such as a phone or hardware token)
3. Something they *are* (referring to biometrics, such as a fingerprint)

Several 2FA methods are in use. Methods such as SMS, TOTP (time-based one-time password), and hardware code generators (such as the RSA SecurID) require the user to enter a single-use code in addition to their password. These codes are either sent to the user via a separate channel or are generated on the fly by the user's device. In commercial and government settings, smart cards are a commonly used second factor, requiring the user to insert an ID badge into a card reader attached to their computer. Online banking systems, particularly in the UK, frequently use variants of hardware code generators and card readers in their 2FA implementations. Companies including Google, Dropbox, and Github have deployed USB hardware tokens (aka security keys), such as YubiKey, internally [18].

Two-factor authentication provides a strong defense against account compromise. The number of recent password database leaks [2] underscores the risk of account compromise. Because users tend to reuse the same username and password across multiple sites [11], password leaks from a single site can lead to a chain-reaction of account compromises as attackers access other accounts with the same credentials [15]. Even if an attacker steals or guesses a user's password, the attacker must compromise the user's phone or steal a physical token to gain access to the account. Thus it is significantly more difficult for a remote attacker to compromise an account protected by a second authentication factor.

Despite the attractive security benefits of 2FA, its impact on the user experience remains unclear. Previous studies on


*USENIX Symposium on Usable Privacy and Security (SOUPS)* 2019.
August 11–13, 2019, Santa Clara, CA, USA.

2FA have produced results which may appear contradictory. While one set of studies [14] [16] [17] [25] concludes that 2FA is completely unusable, others [13] [18] find that some 2FA methods are actually very usable.

It is difficult to draw general conclusions from these prior surveys and studies because of widely-differing conditions. These confounding factors make it very difficult to determine how the different methods compare in terms of usability.

We conducted a two-week, between-subjects usability study of five common 2FA methods with 72 participants, collecting both quantitative and qualitative data. Participants logged into a simulated banking website nearly every day using 2FA and completed an assigned task. Having all the participants experience 2FA within the context of a single application reduces the confounding factors that are usually present when comparing the results of different 2FA methods across usability studies. Participants generally gave high marks to the methods studied, and many expressed an interest in using 2FA to provide more security for their sensitive online accounts.

We purposely ignored setup issues during our initial study to not bias participants toward the day-to-day usability of one of the factors based on a poor setup experience. However, the promising results from the two-week study leave open the question about whether encouraging results for a given factor are incomplete if there is an associated usability hurdle to set up that factor. To gain insight into this question, we conducted a within-subjects laboratory study of the setup process for the five 2FA methods. While a few participants experienced difficulty setting up a hardware token and a one-time password, in general, users found the methods easy to set up.

## 2 Related Work

Previous research has explored the usability of 2FA methods through lab studies and surveys.

### 2.1 Lab Studies

Ace et al. [4] studied the setup and login of four of Google's 2FA methods. They found that participants experienced many failures and found Google's 2FA system hard to use. The order of preference of the four systems reported in their study exactly match the preference ordering of those four systems in our study, but setup results differ significantly. Our differing results may be explained in part because they measured 2FA setup and login with the same participants in a single study. Also, Google changed the setup instructions between their setup study and ours, which may account for our more positive setup results.

Weir et al. [25] compared the usability of three hardware code generator under evaluation by a bank in the UK. Users preferred the system that was most convenient over systems with stronger security. Weir et al. [26] also conducted a lab study of three authentication systems, including SMS and hardware code generator based two-factor systems. Participants were most successful using the SMS-based system.

Lang et al. [18] report on Google's internal deployment of security keys to their employees. They report a long-term reduction in the number of authentication-related support tickets after deploying the hardware keys. Further, they demonstrate a significant reduction in overall authentication time compared to other one-time code based methods.

Das et al. [12] performed two studies measuring both the usability and the acceptability of using the YubiKey (a type of FIDO U2F compliant hardware token) as a second factor in securing a Google account. Employing a think-aloud protocol, they made some recommendations to Yubico (the manufacturer of the YubiKey) based on common points of confusion. After one year, they repeated the study with a new group of users and found that although many of the previous usability concerns had been addressed, many users still did not see much benefit in using the YubiKey. Das et al. postulated that this lack of acceptability was due partly to the lack of awareness of the risks mitigated through using the YubiKey.

Reynolds et al. [21] describe two usability studies of YubiKeys. The study found many usability concerns with the setup process of the YubiKey but found that day-to-day usability was significantly higher. Similar to our study, participants used the YubiKey for several weeks, although we studied the YubiKey in conjunction with several other 2FA methods.

### 2.2 Surveys

Krol et al. [17] conducted interviews with 21 individuals who used two-factor authentication as part of the login process for several UK banks. Participants used a variety of two-factor methods, including card readers, hardware code generators, SMS, phone calls, and smartphone apps that generated single-use codes. Participants particularly disliked hardware code generators; in fact, a few individuals changed banks because of the difficulty of using the tokens. De Cristofaro et al. [13] conducted a Mechanical Turk survey of participants with experience using hardware code generators, one-time codes via SMS and email, and smartphone code generator apps. They found that email or SMS messages were the most commonly used second factor for financial or personal sites, and hardware tokens were the most common for work. Each of the methods received SUS (System Usability Scale) scores in the 'A' range.

Duo is a commercial 2FA product that supports second-factor authentication using a smartphone, phone calls, U2F, and several other methods. Weidman and Grossklags [24] studied the transition from a token-based 2FA system to Duo for employees through a survey at Pennsylvania State University. They found that employees preferred the prior

token-based system to using the Duo app. Some employees' preference was influenced by their dissatisfaction with being required to use personal devices for work. Colnago et al. [10] conducted a large-scale survey of faculty and students at Carnegie Mellon University during a campus-wide deployment of the Duo 2FA system. The results showed that many participants in the survey recognized the security benefits of using 2FA. They also identified usability issues with the deployment of Duo. Differences in perceived usability between users that *voluntarily* adopted 2FA and those that were *required* to adopt 2FA were fairly small, and many participants that were required to use 2FA reported it to be easier to use than they expected.

## 3 Five 2FA Methods

We compare five common 2FA methods: SMS, TOTP, pre-generated-codes, push, and U2F security keys. The differences between our study and the prior work is that we study all five methods in the context of a single simulated web application to reduce the potential for confounding factors and to be able to measure the time to authenticate using each method. We also separate setup and daily use. We are also the first study to include pre-generated codes. This section describes each method and its security properties.

### 3.1 SMS

One of the most widely deployed 2FA methods is SMS. The user is sent a one-time verification code (usually six digits) through a text message to their mobile phone. The broad deployment is partly because most consumers already own a mobile phone capable of receiving text messages—99% of Americans according to a recent Pew study [3]. Potential usability problems may include delayed delivery, lack of cellular service (such as in a foreign country or remote location), and miscopying the code from phone to computer.

SMS-based authentication is vulnerable at several stages. Mobile networks do not encrypt messages while in transit, allowing attackers to conduct man-in-the-middle attacks. Of particular concern, is the well-documented SIM-swapping attack [5, 20] Also, the server (or relying party) must securely store the one-time code while the SMS message is sent, received by the user, and entered back into the site for verification. The code could be salted and hashed to prevent casual theft, but a determined attacker could easily conduct a brute force attack on a stolen hashed code given the relatively small number of codes. Attackers may also steal SMS codes through targeted phishing attacks. Some ways to mitigate these threats are to invalidate a code after a short time window and limit the number of failed attempts to log in with a code.

### 3.2 TOTP

To set up TOTP, the user first synchronizes a secret key generated by the provider to their smartphone, usually by scanning a QR code. The app generates a verification code by combining the secret with a truncated timestamp, hashing the value, and truncating the result to derive the verification code (as with SMS, usually 6 or 7 digits long). The server verifies the user-supplied code using the same method. The advantage of using a TOTP code generator app is that after syncing the secret to the phone, the user does not need to rely on a cellular provider to deliver the one-time codes—eliminating both a potential attack surface and a problem with usability. However, if an attacker steals the TOTP secret from the server or the phone, then the attacker may be able to impersonate the user.

Each code is valid for a set time interval, usually only 30 seconds, after which a new code must be generated. The smartphone and the server must both have a clock that is reasonably in sync. A server accepts tokens for the current 30-second window and the 30-second window just before and after the current window to account for clock drift. Crucially, this means that users may have as little as 30 seconds to enter the code because codes can be generated anytime during the 30-second interval. As with SMS, the verification codes still must be manually keyed in by the user, leaving additional room for user error. According to Pew [3], 77% of Americans own a smartphone, meaning that TOTP is not as broadly deployable to all customer bases as SMS.

TOTP requires a shared secret key between the server and the user's mobile device. This secret must be stored securely, but a one-way hashing mechanism is not useful since the secret is an input to the code-generation and verification process. On the server side, the shared secret could be encrypted using the user's password to prevent casual theft. Assuming secure storage of the shared secret on both the client and server, TOTP has a significant advantage over SMS codes—it does not rely on the insecure mobile network for delivery of the code, thus eliminating an entire attack surface.

### 3.3 Pre-generated Codes

Pre-generated codes are often a backup 2FA authentication method in case the user is unable to access their primary 2FA method. Implementation is straightforward: the service provider generates a list of verification codes and has the user print or write the codes down. The length of the list itself is variable, and the codes are usually around 8 digits long. The codes may be used in any order and must be kept secure by both the server and the user to prevent theft. Because these codes are usually longer than the codes sent through SMS or generated with TOTP, there is additional room for user error when entering the codes. Furthermore, the user must be careful not to lose the medium on which they recorded the

codes.

Printed codes are usually used as a backup authentication mechanism, and must be stored on the server for long periods. Even applying the hashing mechanism discussed for SMS codes, the non-expiring nature of the codes would make them vulnerable to an offline brute-force attack. Although more technically complex to implement, one mitigation against a brute-force attack would be to hash the backup code with the user's password. On the user's side, the printed codes must be stored securely using traditional physical security measures. An open question is how users would prefer to store such backup codes—do users prefer to keep the codes on their person for convenience (perhaps storing the codes in a wallet or purse), or would they prefer to take more stringent security measures to safeguard the codes.

### 3.4 Push

In the push method, the user receives a push notification on their smartphone that allows the user to either "Approve" or "Deny" a login attempt. Push authentication requires Internet access. Google supports this technique (through their "Google prompt"), and it is also available through commercial applications such as Authy OneTouch and DUO Mobile. The advantage of this method is that there is less chance of user error since there are no numbers to copy off a phone screen correctly. We hypothesize that not having to type in numbers, as required by other 2FA methods, is both faster and perceived as more usable by participants.

Push authentication does not explicitly require the storage of a secret key; however, the server must ensure that the push notifications are sent to the correct device, suggesting that some form of two-way verification of the client and server must take place. Additionally, communication between the user's device and the server must be kept secure, such as through the use of TLS. The most prominent push-based authentication methods are proprietary, making it difficult to verify the exact security measures in place and requires implicitly trusting a third party. Push-based authentication has not yet been well-studied by the security community.

### 3.5 U2F Security Keys

Originally developed through a collaboration with Google and Yubico, and now sponsored by the FIDO (Fast IDentity Online) Alliance, Universal 2nd Factor (U2F) is an open standard for authentication using a USB hardware device. To authenticate with a security key, the user must connect the device to their computer and activate the device when prompted by the website.

The U2F standard was designed to be highly secure while still boasting good usability [18]. In contrast to the other four 2FA methods described above, the U2F standard itself is designed to prevent phishing attacks and provide more security and privacy protections than other forms of 2FA. U2F authentication requires that the server store a public key that the user generates at registration time—the secret key never leaves the U2F device. The main risk is that a user might lose their U2F device—but device loss is also a risk with the other four 2FA methods.

## 4 Two-week Study Methodology

We conducted an IRB-approved, two-week study of 2FA at Brigham Young University (BYU). The research objective of the study was to compare the usability of the five common 2FA methods described in Section 3.

### 4.1 Study Design

A total of 72 participants were divided into 6 groups of 12 participants each. Five of the groups were assigned to a specific 2FA method, and the final group was a control group that used only passwords with no second factor. Each participant initially met with a study coordinator to create an account on the study website. During this meeting, the participant was given a list of 12 tasks to complete on the study website over the next two-week period (with no more than one task completed per day). As part of completing each task, each participant would log in to the study website using their assigned authentication mechanism. After the two weeks, participants returned for an exit interview with a study coordinator. Using a combination of authentication event timing data, survey responses, and qualitative data gathered from the exit interviews, we compared the usability of the various authentication methods under test and made some observations and recommendations based on this data.

### 4.2 Banking Website

Our test scenario was that of a participant needing to log in to an online banking interface and complete a task, such as transferring money between accounts or paying a bill online. To support this scenario, we built a simulated online banking interface. supported authentication through either a password alone or a password plus one of the five 2FA methods described previously.

### 4.3 Recruitment

We recruited the 72 participants using flyers posted throughout the BYU campus. Prospective subjects were informed that they would need daily access to an Internet-connected computer with Google Chrome. We required Chrome because it is the only major browser that supports U2F security keys by default. To be eligible for the study, potential participants

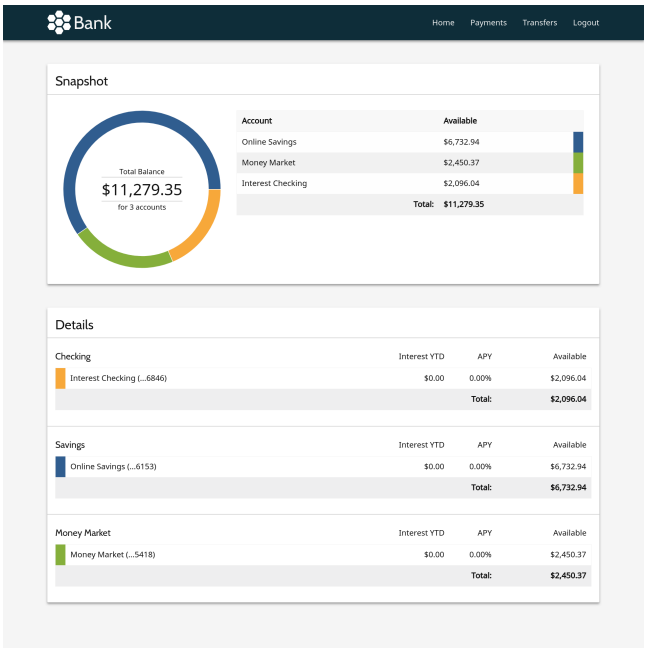



Figure 1: Example of the banking interface we constructed for our study

filled out a short survey to report whether they owned an Android or iOS smartphone, or if they owned a phone able to receive text messages.

Participants were then randomly assigned to a study group. One participant did not own a smartphone and was randomly assigned to a study group that did not require the use of a smartphone. Once a group reached 12 participants, we removed it from the pool of potential groups to which a participant could be assigned.

## 4.4 Demographics

We had a slightly higher number of female participants (38; 55%) as compared to male participants (31; 45%) in our study. Participants were largely young adults: 18–19 years (3; 4%), 20–29 (61; 88%), and 30–39 (5; 7%). Over two-thirds of the participants (49; 71%) had completed some college but had not yet completed a degree. Participants self-reported their level of computer expertise: far above average (13; 19%), somewhat above average (28; 41%), average (25; 36%), and somewhat below average (3; 4%).

## 4.5 Setup and Initial Meeting

Participants scheduled an initial appointment to meet with a study coordinator. During the initial meeting, the study coordinator assisted them in setting up an account on the online banking interface. We allowed participants to choose their username and password, with the only restriction being that the password had to be at least eight characters long.

If the participant belonged to one of the study groups using a second-factor method, the coordinator would also help them configure 2FA on their account for the study website. Depending on the study group, this included helping the participant install any necessary apps (Authy for push, Google Authenticator for TOTP), verifying their phone number, issuing the participant a U2F device (the YubiKey NEO), or printing the backup codes. Finally, the study coordinator assisted the participant in completing the first listed task during the initial meeting, leaving the participant with 11 tasks to complete on their own.

For this study, we purposely chose to focus only on the day-to-day use of 2FA methods and not confound those results with any negative issues arising from the usability of the setup process. Recent papers have studied 2FA setup of YubiKeys [12, 21], for instance, and argue that researchers should examine setup and day-to-day use independently. If day-to-day use is acceptable and promising to users, this can motivate more energy to address problematic setup procedures.

## 4.6 Two-week Task Completion Period

Over the next two weeks, participants were asked to complete no more than one task per day in the order given on their task list. To complete each task, the participant would need to visit our online banking website and log in with their previously selected username and password. Except for the control group using only a username-password pair, the participant would also authenticate using their assigned second-factor method for each login. After logging in, the participant would go to either the "Payments" or "Transfers" page and complete the banking component of the task. The purpose of having participants complete the banking-related task after logging in (as opposed to merely having the individual log in and do nothing) was to encourage the user to act more naturally during the login process and make the simulation more realistic—most real-world users do not authenticate for amusement; instead authentication is a means to an end.

## 4.7 Exit Interview

Participants reported back for an exit interview with a study coordinator after the two weeks. The coordinator first had the participant take a brief survey to gather a small amount of demographic data. Participants also completed a SUS (System Usability Scale) assessment of the website as a whole, and for the authentication method they had used during the study. Following this, the coordinator conducted a semi-structured interview with the participant to gather additional information about how the participant felt about the website overall as well as the login process. In particular, we asked participants questions about their overall online security posture to better

Table 1: Repeated measures correlation (rmcorr) between amount of time participating in study versus amount of time to authenticate.

| 2FA Method | p-value | r | df | 95% confidence interval |
|---|---|---|---|---|
| SMS | 0.280 | -0.097 | 124 | (-0.269, 0.081) |
| TOTP | 0.586 | -0.049 | 122 | (-0.225, 0.129) |
| Push | 0.029 | -0.204 | 113 | (-0.374, -0.020) |
| U2F | <0.003 | -0.269 | 118 | (-0.429, -0.093) |
| Codes | 0.426 | -0.076 | 110 | (-0.260, 0.113) |

understand their background and feelings about online security. With the consent of each participant, we recorded the audio of each interview. Two coders listened to the recordings and coded each interview, discussing each response until reaching agreement. Common themes identified from the recordings are discussed in section 5.2.

## 4.8 Compensation

Participants were compensated a maximum of 25 USD after their participation in the study according to a tiered compensation structure based on the total number of tasks completed through the banking interface.

# 5 Two-week Study Results

## 5.1 Quantitative Results

### 5.1.1 Timing Data

We measured both the time for the password login and the time for the 2FA on the server side based on events sent from the client. Password timing began when the page initially loaded and ended when the user submitted a password. 2FA timing began when the 2FA prompt was loaded and ended when the 2FA was verified (or rejected). We recorded timestamps on the server since each client may have a slightly different clock. By comparing adjacent timestamp events, we were able to compute the overall login time. It is possible that users spent time obtaining their 2FA device before accessing the login page, which is not accounted for in the timing data.

**Individual Learnability** We computed the correlation between the amount of time an individual had been in the study and the amount of time it took them to authenticate. We used the repeated measures correlation (rmcorr) technique described by Bakdash and Marusich [6] to estimate the common regression slope for each 2FA method in our study. We hypothesized that participants would get faster over time as they became more familiar with the 2FA method. We found statistically significant ($p < 0.05$) support for this hypothesis for both push notifications and U2F security keys (see Table 1).

Table 2: Authentication Time (seconds), Summary Statistics

| Authentication Method | Q1 | Median | Mean | Q3 |
|---|---|---|---|---|
| Codes | 11.3 | 17.2 | 28.0 | 25.4 |
| Push | 8.4 | 11.8 | 16.1 | 17.6 |
| SMS | 13.0 | 16.6 | 18.5 | 22.1 |
| TOTP | 10.7 | 15.1 | 23.9 | 23.3 |
| U2F | 4.5 | 9.1 | 13.0 | 16.3 |

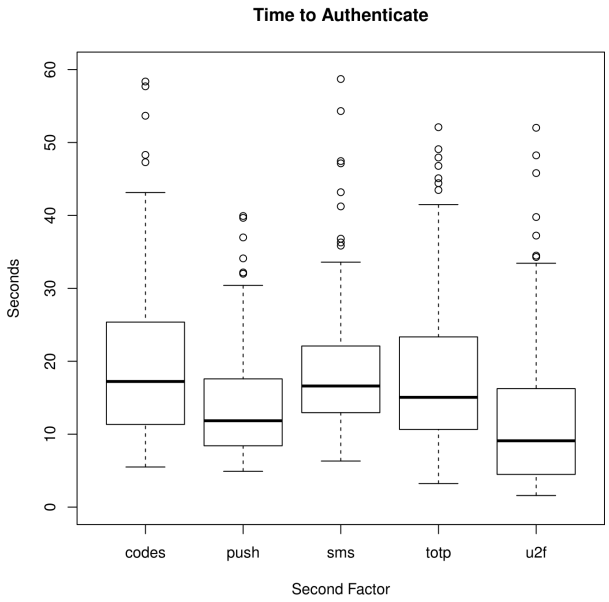



Figure 2: Time to authenticate for five 2FA methods

**Comparison of 2FA Authentication Times** We applied a Kruskal-Wallis one-way analysis of variance and found there was a significant difference ($p < 0.001$, $\alpha = 0.05$) in the median authentication time between the methods. We did not include the time that it took the user to enter their password; the observed authentication times reported here include only the time to get through the second-factor authentication step. The security key (U2F) devices had the fastest median authentication time, followed by push notifications. These timing results are summarized in Table 2 and Figure 2.

### 5.1.2 Usability Survey Rankings

We administered two SUS surveys to participants at the beginning of each exit interview session. The first survey addressed

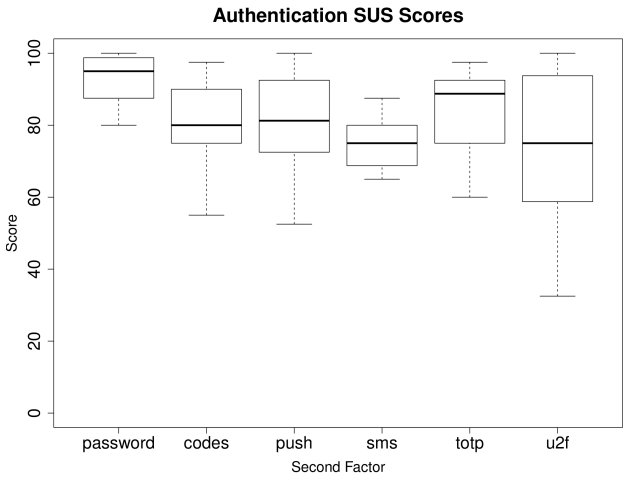



Figure 3: SUS scores for five 2FA methods.

the usability of the banking website as a whole, and the second addressed only the usability of the login system. The purpose of administering two surveys was to determine how large an impact the banking website itself had on the participants' feelings about the authentication method. Additionally, we felt that participants would be more accurate with their opinions about the 2FA method if we had first given them an opportunity to both consider and express their feelings about the system as a whole; had we only given a SUS survey on the authentication method we felt participants would be more likely to (incorrectly) report their feelings about unrelated website features.

The SUS scores for the authentication methods are summarized in Figure 3 and Table 3. We performed a Kruskal-Wallis one-way analysis of variance and determined that the authentication method used was a statistically significant ($p = 0.02579$, $\alpha = 0.05$) predictor of the median SUS score for the 2FA method. We also computed the value of $\rho = 0.7576$ for Spearman's rank correlation coefficient and confirmed that there was a significant ($p < 0.001$) correlation between the overall website SUS scores and the SUS scores of the individual authentication methods. Passwords with no second factor had the highest median SUS score, with a median score of 95, followed by TOTP (via Google Authenticator) which had a median SUS score of 88.75.

## 5.2 Qualitative Results

### 5.2.1 Security and Inconvenience

We asked participants whether logging in with a second verification step felt more secure. Most participants did feel more secure, although 3 of 12 participants that used the printed backup codes did not feel like the codes added any additional security to the method.

Table 3: SUS Scores for each Method, Summary Statistics

| Authentication Method | Q1 | Median | Mean | Q3 |
|---|---|---|---|---|
| Password | 87.5 | 95.0 | 92.5 | 98.8 |
| Codes | 75.0 | 80.0 | 80.2 | 90.0 |
| Push | 72.5 | 81.3 | 81.0 | 92.5 |
| SMS | 68.8 | 75.0 | 75.0 | 80.0 |
| TOTP | 75.0 | 88.8 | 83.1 | 92.5 |
| U2F | 61.9 | 75.0 | 73.1 | 93.1 |

> **P6:** *"I felt like the codes didn't accomplish anything, because that's just more passwords—anyone could guess them."*

We also asked participants if the additional security would be worth the additional login time or inconvenience they might face when using the second-factor method. Several people (20; 29%) said the extra security was definitely worth the tradeoff, and an additional group (25; 36%) said that they would be willing to use 2FA depending on the importance of the account.

> **P25:** *"In my opinion, it may be a little obsessive for everything, but for banking it's something that I actually do want some authentication. I almost wish that it was a requirement that the bank said, oh here set [two-factor authentication] up. Because now that I think about it, I don't know how to set up 2FA with my bank. If it were an option I would definitely use 2FA."*

> **P33:** *"It was pretty quick, so that was good; I didn't feel like I had to jump through a lot of hoops. I can imagine it being nice having an extra wall of security if it's your bank information so that even if somebody else gets your password, it's not like they're going to be able to hack into your account because they don't have the [security key]."*

Some participants were particularly concerned about the centrality and importance of their email account, particularly considering the potentially large amount of sensitive data stored there. For example, one participant reported they had already turned on 2FA for their Gmail account to gain extra protection:

> **P24:** *"I use my email for everything, and so I thought it wouldn't hurt to have some extra security. The thought of someone hacking into [my account] and having everything vulnerable... better to be safe than sorry."*

Other participants (9; 13%) expressly stated that they would not be willing to use 2FA to gain additional security because the inconvenience was too high.

> **P37:** *"I don't know how much my level of convenience and my need for level of security would balance out because for me having something that is convenient and is at hand is almost more important than having something that is more secure... I know if people hack your credit cards, then the bank will take care of that and get the money back and so having that extra security makes me care less about having a second factor."*

#### 5.2.2 Availability of Second-factor Device

Each participant in the study in one of the 2FA groups was required to use something external to their computer to login, whether it be the sheet of paper with printed codes, a YubiKey, or their phone. Many participants (24; 35%) mentioned not having their second factor immediately available to them when they needed to log in.

> **P8:** *"I don't always have my phone on me, and so if I'm doing something on the computer, I'm usually doing homework, so I actually try to keep my phone away from me."*

> **P42:** *"Honestly, once I'm home I kind of just set my phone down and forget where I put it sometimes, so that was a little bit hard ... I needed to go find my phone and pull up the app."*

#### 5.2.3 TOTP Timeout

Although the participants using TOTP (via the Google Authenticator app) were overall very positive about their experience, 8 of 12 participants mentioned that they had problems entering the six-digit verification code before it timed out.

> **P30:** *"I have to type in these numbers so fast or else it's going to go away."*

#### 5.2.4 Likelihood of Account Compromise

Participants expressed a wide spectrum of views on how much value they placed on their online accounts. Some participants (9; 13%) felt that they had nothing to protect and would therefore not be a target of criminals.

> **P5:** *"I guess maybe because it's that I don't have anything to protect... I'm at a stage in my life where nothing I own is that valuable and none of my information is that wanted that it makes a difference."*

Table 4: Account Compromise and Inconvenience

| 2FA worth the inconvenience? | Hacked | Not Hacked |
|---|---|---|
| Definitely | 11 | 9 |
| Sometimes | 6 | 19 |
| Never | 4 | 5 |

> **P8:** *"I mean, you hear a lot about stuff being broken into; I just don't think I have anything that people would want to take from me, so I think that's why I haven't been very worried about it."*

> **P30:** *"I don't have a lot of money in my accounts right now, so if someone stole my money, that would be bad, but it's not enough that it would be the end of the world if I lost all my money— I don't feel like I'm a target for someone to steal my stuff. I can imagine in the future if I had a huge retirement fund or something then I would want that to be more secure."*

#### 5.2.5 Prior Compromise vs. 2FA Inconvenience

We asked each participant in this study whether any of their online accounts had ever been compromised. Several participants (26; 38%) described experiences with remote attackers taking over their online accounts, and a few people (7; 10%) mentioned that someone they know has had one of their online accounts hacked. Although not directly a form of online account compromise, a few participants also mentioned experiences with financial theft from having their credit or debit card number stolen or having their bank account credentials stolen. Others mentioned having their personal information stolen as part of one or more data breach events, including the highly publicized Equifax compromise of millions of individuals' personally identifying information [7]. When asked how they noticed that their account was compromised, most participants said they received an email indicating a new login from a suspicious location. We hypothesized that participants with previous experience having an account compromised would be more likely to feel that using a second factor was worth any extra inconvenience. Using data extracted from coding the interviews (see Table 4), we used Pearson's chi-squared test with two degrees of freedom to test the dependence of these variables. Not all participants expressly talked about both of these variables; thus we analyzed only participants for which we had coded data for both variables.

We observed no statistically significant relationship between a participant's previous history with account compromise and whether they felt that two-factor authentication was worth the inconvenience ($\chi = 4.6332$, $p = 0.0098$, $\alpha = 0.05$). One limitation of this analysis is that it does not consider the exact nature of the previous account compromise (such

as whether a financial loss was involved). However, we do note that numerous individuals independently stated that using 2FA would be worth the inconvenience at least some of the time, particularly for financial accounts.

## 5.3 Discussion

In this section, we further highlight some of the most interesting results of our study and discuss their meaning in the context of usable 2FA.

### 5.3.1 Relationship between Authentication Time and Usability

Although both push-based authentication and the U2F security keys had faster median authentication times, neither of these methods received the highest median SUS score. Conversely, TOTP was the highest scoring second-factor method we tested but had a median authentication time that was slower than either push or U2F. From our exit interviews, we identified some explanations for this result. First, some participants receiving push requests through Authy did not always receive the authentication request in their notification area and instead had to open the app and approve the request manually. It was unclear whether this was a bug in the Authy or the result of notification configuration on some participants' phones. Several U2F participants using both Windows and Mac operating systems reported a variety of minor troubles getting the YubiKey to work with their computers (possibly because they plugged it in the wrong direction). However, other participants reported no problems using the YubiKey. Ultimately, participants using TOTP reported liking the relative simplicity of the Google Authenticator app. The app functioned very similarly to SMS, a 2FA method with which many participants were already familiar while not requiring them to always have cellular service.

We believe that the minor issues encountered by participants using the Authy app and the YubiKey likely explains most of the lower scores they received. That said, no authentication method we tested received a poor usability rating, suggesting that, although there is a noticeable impact on usability from requiring 2FA, the presence of 2FA itself does not doom the method as a whole to poor usability.

### 5.3.2 Remember Me?

A novel aspect of our study is that participants used their second factor repeatedly for two weeks instead of using it just once in a laboratory setting. We purposely did not provide a "Remember Me" option, thus requiring participants in the non-control groups to use their second factor every day. We believe that some of the usability impacts of needing a second factor could be mitigated by only requiring the second factor on new computers or after logging out. Requiring less frequent 2FA login would provide a similar level of protection against remote attackers while mostly allowing users unfettered access to their accounts. Some systems allow access for a limited amount of time (30 days, for instance) without requiring a second factor on the same machine. Participants with previous experience using such systems (typically for a university login system) made some remarks to the effect that they were never quite sure when the second factor would be required. One solution to this problem would be to have a small count-down displayed to the user telling them how many days were left until they would need to provide their second factor again to avoid the "ambush" effect described by Sasse et al. [22]. Further research needs to be done to determine the right balance of when to ask the user for the second factor again when they have already been logged in previously on the same machine.

### 5.3.3 Positive Feedback

Given the weak usability results of previous 2FA studies, we expected an overall poor usability response. During the exit interviews, we were surprised at the number of participants that reported an overall positive experience using 2FA. Many participants wanted to use 2FA for some of their actual online accounts but were either unaware it was an option or were unsure how to configure it.

### 5.3.4 Differentiating Between High-value and Low-value Accounts

Although participants generally tended to care less about the security of their social media accounts, many expressed concern about the security of their banking and financial accounts. There were mixed feelings about frequently used accounts like email accounts, however, particularly in balancing whether it would be worth using 2FA for such accounts. Participants generally agreed that they did not want to be required to use their second factor to log in to their email account from a known computer. Other participants felt they had no confidential information in their email, and that having a second factor would not be worth the extra login step. In general, the higher the perceived value of the account, the more likely the participant was to be willing to use 2FA for the account.

## 5.4 Limitations

Our study has several limitations. First, the participants were not asked about their prior use of 2FA. A user assigned to a second-factor they were already familiar with could bias the results. Second, the participants were university students that were younger and more technically savvy than the general population. The students are also more likely to have fewer material assets to be concerned with, as discussed in the qualitative results. Third, we deliberately chose not to have the participants setup the 2FA mechanism on their own so that a

poor setup experience would not negatively bias day-to-day usability. This decision means the day-to-day usability results could be biased more positively compared to users that will have to setup and use 2FA. Fourth, because we wanted to capture authentication timing data, we were unable to have participants use a real banking system or an existing online account; this may have altered their behavior. Fifth, participants were required to use 2FA for every authentication attempt, which may have caused them to acclimate to using 2FA more quickly than would be seen if 2FA was required only on new machines. Sixth, participants' discussions of the necessity of 2FA and online security may have been different had we mocked our website as a social media site. Finally, with only 12 participants in each study group, we may not have reached saturation in the qualitative data that was gathered. Even if we had reached saturation, the limited demographics of the study still warrant further studies with a broader population.

# 6 Setup Study Methodology

We purposely ignored the setup phase during our two-week study to avoid having a poor setup experience negatively bias participant's evaluation of the day-to-day usability of one of the factors. However, the promising results from the study beg the question about whether the results are incomplete and miss an important associated usability hurdle to set up that factor. To gain insight into this question, we conducted an IRB-approved, within-subjects laboratory study comparing the usability of the setup phase for the five 2FA methods. Based on our initial review of the setup process on some popular websites, we did not expect that there would be significant usability issues for setting up the five 2FA methods.

## 6.1 Study Design

Each participant was tasked with setting up the five 2FA methods from a desktop computer using a provided Google account. We chose to test the methods on Google because it supports all five 2FA methods and is an industry leader in supporting 2FA for its customers and employees. The setup for Google security keys has been studied previously, and improvements have been made based on those results [12, 21]. Our goal was to observe the general usability of the setup process and not focus on provider-specific details since we did not compare the setup between multiple providers.

Participants were provided with an Android phone and a YubiKey NEO for methods requiring a physical device. Testing for every possible ordering of setting up the five methods requires 120 treatments. To reduce the time and cost of our study, we created an incomplete counterbalanced measure designed to mitigate biases due to the order participants set up each of the 2FA methods. We used two five-by-five balanced Latin squares to generate 10 different orderings of the setup methods to counterbalance sequential effects caused by ordering [19]. Each of the 10 orderings was completed 3 times during the study. After each attempt to set up the second factor, participants were asked to complete the Single Ease Question (SEQ) to measure the difficulty of each task. The SEQ is a standard usability questionnaire with a single question ("Overall, how difficult or easy was the task to complete?") rated on a 7-point scale. Although it contains only one question, the SEQ has been found to perform reliably [23]. We chose SEQ to avoid survey fatigue since participants were asked to rate five different methods. We used timing data and SEQ responses to compare the setup usability for the five methods.

We posted flyers on the BYU campus to recruit 30 participants who were familiar with Google accounts and Android phones. As each participant met with a study coordinator, they first signed a consent form. Participants were compensated 10 USD at the conclusion of the study. We assigned participants to the ten Latin square orderings in a round-robin fashion.

## 6.2 Setup Process

The coordinator provided the participant with an Android phone, a YubiKey, and an information sheet that listed the cellphone number and lockscreen passcode. We made an audio recording of the verbal comments of each participant along with a video recording of the computer screen.

Google does not allow backup codes, push notifications, or TOTP to be set up without first setting up SMS or U2F. In order to test each method independently, we used one Google account for setting up SMS, and a separate account for the other four options. Study coordinators navigated to Google's 2FA setup page on a Chrome browser and then instructed participants in what order to set up the five second factors. The coordinators also navigated between the two Google accounts before and after the participant setup SMS. After the participant completed the task or was unable to finish the setup, the coordinator prompted the participant to complete the SEQ. Coordinators did not assist participants in setting up any of the second factors.

The following is a brief summary of each setup task.

**SMS.** Participants were asked to type in a phone number. Google sends a confirmation text containing a six-digit code to the provided number. The participant completes setup by entering the code on Google's webpage.

**TOTP.** Participants were provided with an Android phone without Google's Authenticator app installed. We wanted participants to download the app as a part of the setup because we assumed that the typical Google user would not have the app already installed on their phone. The phone was set up with the Play Store app on the home page for easy accessibility. Google instructed participants to install the app using the Play Store, and then to scan a Quick Response (QR) code shown on the webpage. After scanning the QR code, participants completed the setup by entering a six-digit code from the

Authenticator app into the webpage.

**Pre-generated codes.** Google autogenerates 10 backup codes upon request. Participants were not required to print or download these codes but were asked by the coordinator how they would store these codes if they were using their own Google account. Some participants shared that they would choose to take a photo of the codes using the camera on their phone, while others said they would write down the codes and keep them in a safe place. For timing data, we measured from the time the participant began the task to the time the backup codes were displayed on the screen. Even though we asked participants how they would store the backup codes, we did not include the time taken to store codes in the setup time for backup codes since the time to store the codes varies widely depending on the storage method chosen.

**Push.** Push notifications require that the phone is signed in to the user's Google account. The phone provided to participants was already signed in, based on the assumption that the typical Google user would already be signed in to their Google account on their phone. When a phone is online, has screen locking enabled, and is connected to the Google account, Google sends a push notification that can be approved by unlocking the phone and tapping "Yes" on the notification.

**U2F Security Key.** We provided participants with a YubiKey NEO. Google directed participants to insert the security key into an open USB port, and then to tap the gold button on the key. Before the device could be recognized, participants were required to dismiss an alert from the browser asking for permission to see the U2F device's make and model. Whether or not a user allows or denies this request, the U2F device is registered and optionally given a name. Since this is optional, we excluded the time taken to name the device.

# 7 Setup Study Results

## 7.1 Timing Data

We reviewed the video screen recordings to measure the setup time for each method. Time was measured in seconds from when the participant began the setup task to the time Google notified the participant that the setup had been successful. The cases where the participant failed to complete the setup are not included in the timing analysis. Setup failure occurred twice with the U2F device and twice with the TOTP application. A summary of our results is shown in Table 5 and Figure 4.

As expected, backup codes had the fastest setup time because all that was involved was clicking the webpage button to generate the codes. However, backup codes had the longest mean authentication time in the day-to-day study, followed by push notifications and SMS messaging. While U2F devices had the fastest mean authentication time in our day-to-day study, they had the second slowest mean setup time. TOTP had the slowest mean setup time.

Table 5: Setup Time (in seconds), Summary Statistics

| Authentication Method | Q1 | Median | Mean | Q3 |
|---|---|---|---|---|
| Codes | 1.0 | 1.0 | 2.2 | 2.0 |
| Push | 16.0 | 23.5 | 27.3 | 33.0 |
| SMS | 27.5 | 32.0 | 34.5 | 40.0 |
| TOTP | 73.3 | 84.0 | 109.6 | 120.0 |
| U2F | 31.8 | 44.0 | 57.8 | 67.8 |

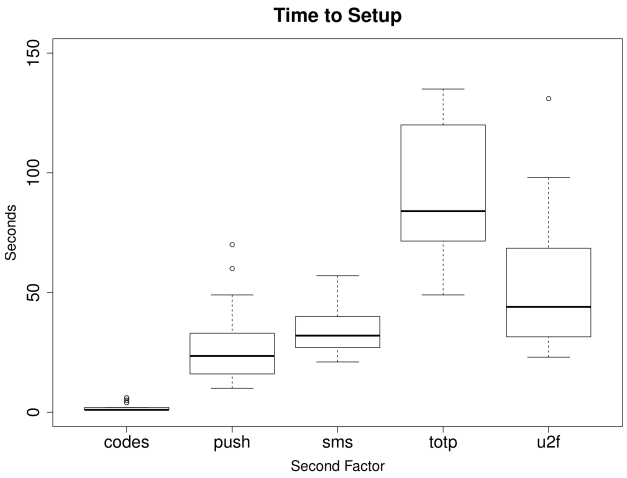



Figure 4: Setup time for five 2FA methods.

## 7.2 SEQ Scores

Participants answered the SEQ after completing (or being unable to finish) each 2FA method. Mean SEQ scores are shown in Table 6 and the distribution of all SEQ scores is shown in Figure 5. With the exception of backup codes, the ranking of best SEQ score to worst corresponds with the ranking of time to set up, i.e. the faster the setup, the higher the mean SEQ score. We were surprised that backup codes received a lower ranking since setup involved nothing more than pushing a button. We hypothesize that a participants' perceptions about the day-to-day usability of the 2FA method influenced their SEQ score even though they were instructed to rate only the usability of the setup task.

TOTP setup received the lowest mean SEQ score of the five methods. The low score is in stark contrast with the two-week study, where TOTP received the highest mean SUS score of all five 2FA methods. Users may have been more unfamiliar with setting up TOTP then they were with the setup of more common methods, such as SMS. However, once users have TOTP authentication successfully enabled, they may find it to be more usable than other 2FA methods they may have traditionally relied on.

Table 6: Mean SEQ Scores

| Push | SMS | Codes | U2F | TOTP |
|---|---|---|---|---|
| 6.7 | 6.2 | 5.9 | 4.7 | 4.5 |

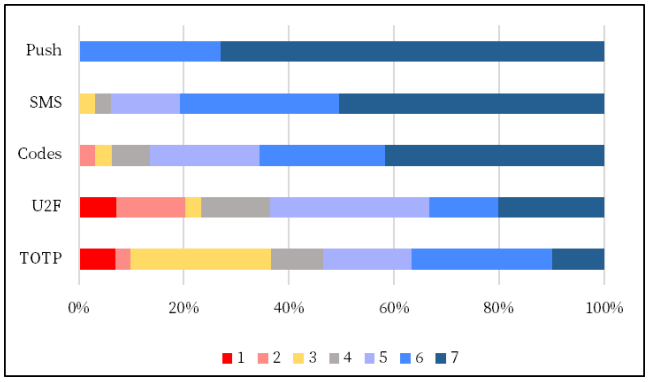



Figure 5: SEQ scores for five 2FA methods.

## 7.3 Discussion

Our study suggests that when 2FA setup can be implemented well, users generally find it easy to accomplish. Each of the five second factors had a mean score closer to the "easy" side than the "difficult" side. This is notable considering that study coordinators provided no assistance during setup, and many participants were required to set up second factors that were unfamiliar to them (such as the U2F device or the TOTP generator). SMS authentication is one of the most common forms of 2FA, and familiarity with using SMS as a second factor likely influenced its SEQ score.

Setup failure occurred twice with TOTP and twice with U2F. Both failures for TOTP happened when the participant immediately attempted to scan the QR code with the phone's camera, instead of downloading the Authenticator app to scan the code. An additional two participants initially tried to scan the QR code with the phone's camera but realized their mistake and successfully completed setup after downloading the app. The failures for U2F both occurred when the participant did not notice the browser alert requesting permission to see the U2F device's make and model. Google does not require the make or model to authenticate the device, so the U2F device would be registered whether or not the user allowed or denied the browser's request. However, participants who did not notice the alert at all were not able to complete the setup.

Based on our observations, we present two recommendations for reducing setup failures on Google accounts. First, users may be less likely to skip over installing the Authenticator app if the installation instructions were on a prompt separate from the QR code. Second, because the U2F browser alert occurs on many of the browsers that support U2F (including Chrome, Opera, and Firefox), 2FA-providers should notify users about the alert during the setup process. Yubico does this on their support page: "Touch the YubiKey when prompted, and if asked, allow it to see the make and model of the device" [1].

## 7.4 Limitations

Participants from our study were recruited at a university, and our results may not be generalizable to the general population. We tested setup on a desktop computer, and the setup experience may be different using a phone as the primary computing platform. Our timing data for backup codes did not include the time taken to store codes. Timing data and SEQ scores may have been negatively impacted by our participants' unfamiliarity with the provided phone. If participants had used a personal phone, they likely would have been able to perform tasks requiring a phone more quickly (e.g., entering the phone number, or unlocking the phone). Although our study did not focus on provider-specific details, Google's implementation of 2FA setup inevitably influenced user's perceptions.

# 8 Conclusion

We conducted a user study to evaluate the day-to-day usability of multiple 2FA methods by having participants log in to a simulated banking website nearly every day for two weeks and completing an assigned banking task. Having all the participants experience a 2FA method within the context of a single application reduces the confounding factors that are usually present when comparing the results of different 2FA methods across usability studies.

Participants generally gave high marks to the methods studied, and many expressed an interest in using 2FA for their sensitive online accounts. However, about one-third of the participants reported an instance of not having their second-factor device immediately available when they needed it.

There are several lessons learned from our two-week study. Participants using push notifications and U2F security keys decreased their login time as they gained experience with the method. Two-thirds of the participants using TOTP (via the Google Authenticator app) had problems entering the six-digit code before it timed out. Approximately 25% of the participants using printed backup codes did not feel like the codes added any additional security to the system—it seemed like just another password that an attacker could compromise.

We also compared the usability of the setup phase for each of the five 2FA methods. While a few participants experienced difficulty setting up U2F and TOTP as second factors, in general, users found the methods easy to set up. Together, our two studies show that well-implemented 2FA methods may be set up and used daily without major difficulty.

## Acknowledgments

The authors thank the reviewers for their helpful feedback. This material is based in part on work supported by the National Science Foundation under Grant No. CNS-1816929.

## References

[1] How to confirm your Yubico device is genuine with U2F. *Using Your YubiKey with Authenticator Codes : Yubico Support*.

[2] Data Breach Investigations Report, 2017.

[3] Mobile Fact Sheet, Jan 2017.

[4] Claudia Acemyan, Philip Kortum, Jeffrey Xiong, and Dan Wallach. 2fa might be secure, but it's not usable: A summative usability assessment of google's two-factor authentication (2fa) methods, 2018.

[5] Nathanael Andrews. "can i get your digits?": Illegal acquisition of wireless phone numbers for sim-swap attacks and wireless provider liability. *Northwestern Journal of Technology and Intellectual Property*, 16(2):78–106, Nov 2018.

[6] Jonathan Z Bakdash and Laura R Marusich. Repeated Measures Correlation. *Frontiers in Psychology*, 8:456, 2017.

[7] Tara Siegel Bernard, Tiffany Hsu, Nicole Perlroth, and Ron Lieber. Equifax Says Cyberattack May Have Affected 143 Million in the U.S., September 2017.

[8] Joseph Bonneau, Cormac Herley, Paul C Van Oorschot, and Frank Stajano. The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes. In *2012 IEEE Symposium on Security and Privacy (SP)*, pages 553–567. IEEE, 2012.

[9] Joseph Bonneau and Sören Preibusch. The Password Thicket: Technical and Market Failures in Human Authentication on the Web. In *The Ninth Workshop on the Economics of Information Security (WEIS)*, 2010.

[10] Jessica Colnago, Summer Devlin, Maggie Oates, Chelse Swoopes, Lujo Bauer, Lorrie Cranor, and Nicolas Christin. "It's not actually that horrible": Exploring Adoption of Two-Factor Authentication at a University. In *2018 CHI Conference on Human Factors in Computing Systems*, page 456. ACM, 2018.

[11] Anupam Das, Joseph Bonneau, Matthew Caesar, Nikita Borisov, and XiaoFeng Wang. The Tangled Web of Password Reuse. In *Network and Distributed System Security (NDSS)*, volume 14, pages 23–26, 2014.

[12] Sanchari Das, Andrew Dingman, and L Jean Camp. Why Johnny Doesn't Use Two Factor: A Two-Phase Usability Study of the FIDO U2F Security Key. In *2018 International Conference on Financial Cryptography and Data Security (FC)*, 2018.

[13] Emiliano De Cristofaro, Honglu Du, Julien Freudiger, and Greg Norcie. A Comparative Usability Study of Two-Factor Authentication. In *Workshop on Usable Security (USEC)*, 2014.

[14] Nancie Gunson, Diarmid Marshall, Hazel Morton, and Mervyn Jack. User Perceptions of Security and Usability of Single-factor and Two-factor Authentication in Automated Telephone Banking. *Computers & Security*, 30(4):208–220, 2011.

[15] Blake Ives, Kenneth R Walsh, and Helmut Schneider. The Domino Effect of Password Reuse. *Communications of the ACM*, 47(4):75–78, 2004.

[16] Mike Just and David Aspinall. On the Security and Usability of Dual Credential Authentication in UK Online Banking. In *International Conference for Internet Technology And Secured Transactions (ICITST)*, pages 259–264. IEEE, 2012.

[17] Kat Krol, Eleni Philippou, Emiliano De Cristofaro, and M Angela Sasse. 'They brought in the horrible key ring thing!' Analysing the Usability of Two-Factor Authentication in UK Online Banking. *Workshop on Usable Security (USEC)*, 2015.

[18] Juan Lang, Alexei Czeskis, Dirk Balfanz, Marius Schilder, and Sampath Srinivas. Security Keys: Practical Cryptographic Second Factors for the Modern Web. In *International Conference on Financial Cryptography and Data Security (FC)*, pages 422–440. Springer, 2016.

[19] James R. Lewis. Pairs of latin squares to counterbalance sequential effects and pairing of conditions and stimuli. *Proceedings of the Human Factors Society Annual Meeting*, 33(18):1223–1227, 1989.

[20] Collin Mulliner, Ravishankar Borgaonkar, Patrick Stewin, and Jean-Pierre Seifert. Sms-based one-time passwords: Attacks and defense. In Konrad Rieck, Patrick Stewin, and Jean-Pierre Seifert, editors, *Detection of Intrusions and Malware, and Vulnerability Assessment*, pages 150–159, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[21] Reynolds, Joshua and Smith, Trevor and Reese, Ken and Dickinson, Luke and Ruoti, Scott and Seamons, Kent. A Tale of Two Studies: The Best and Worst of YubiKey Usability. In *2018 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2018.

[22] Martina Angela Sasse, Sacha Brostoff, and Dirk Weirich. Transforming the 'Weakest Link'—A Human/Computer Interaction Approach to Usable and Effective Security. *BT Technology Journal*, 19(3):122–131, 2001.

[23] Jeff Sauro and Joseph S. Dumas. Comparison of three one-question, post-task usability questionnaires. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 1599–1608, New York, NY, USA, 2009. ACM.

[24] Jake Weidman and Jens Grossklags. I like it, but i hate it: Employee perceptions towards an institutional transition to byod second-factor authentication. In *Proceedings of the 33rd Annual Computer Security Applications Conference*, ACSAC 2017, pages 212–224, New York, NY, USA, 2017. ACM.

[25] Catherine S Weir, Gary Douglas, Martin Carruthers, and Mervyn Jack. User Perceptions of Security, Convenience and Usability for Ebanking Authentication Tokens. *Computers & Security*, 28(1):47–62, 2009.

[26] Catherine S Weir, Gary Douglas, Tim Richardson, and Mervyn Jack. Usable security: User Preferences for Authentication Methods in eBanking and the Effects of Experience. *Interacting with Computers*, 22(3):153–164, 2010.

# Personal Information Leakage by Abusing the GDPR "Right of Access"

Mariano Di Martino[1], Pieter Robyns[1], Winnie Weyts[2], Peter Quax[1,3],
Wim Lamotte[1], and Ken Andries[2,4]
[1] *Hasselt University/tUL, Expertise Centre for Digital Media*
[2] *Hasselt University - Law Faculty*
[3] *Flanders Make*
[4] *Attorney at the Brussels Bar*
{mariano.dimartino,pieter.robyns,peter.quax,wim.lamotte,ken.andries}@uhasselt.be
winnie.weyts@student.uhasselt.be

## Abstract

The General Data Protection Regulation (GDPR) "Right of Access" grants (European) natural persons the right to request and access all their personal data that is being processed by a given organization. Verifying the identity of the requester is an important aspect of this process, since it is essential to prevent data leaks to unauthorized third parties (e.g. criminals). In this paper, we evaluate the verification process as implemented by 55 organizations from the domains of finances, entertainment, retail and others. To this end, we attempt to impersonate targeted individuals who have their data processed by these organizations, using only forged or publicly available information extracted from social media and alike. We show that policies and practices regarding the handling of GDPR data requests vary significantly between organizations and can often be manipulated using social engineering techniques. For 15 out of the 55 organizations, we were successfully able to impersonate a subject and obtained full access to their personal data. The leaked personal data contained a wide variety of sensitive information, including financial transactions, website visits and physical location history. Finally, we also suggest a number of practical policy improvements that can be implemented by organizations in order to minimize the risk of personal information leakage to unauthorized third parties.

## 1 Introduction

On the 27th of April 2016, the European Parliament and the Council of the European Union enacted Regulation 2016/679 on "the protection of natural persons with regard to the processing of personal data and on the free movement of such data" [2]. This regulation, commonly referred to as the General Data Protection Regulation (GDPR), supersedes Directive 95/46/EC and provides a number of additional benefits to natural persons (data subjects) when their data is processed by third parties (data controllers). One such example is the "Right of Access", which allows the data subject (DS) to request whether and which personal data concerning him or her is being processed by the data controller (DC) [2, Art. 15].

As of 25 May 2018, the GDPR became enforceable, meaning non-compliant DCs could face a fine of up to 20 million euros or 4% of the annual worldwide turnover of the preceding financial year, depending on the nature of the infringement [2, Art. 83]. This means that by now, DCs should have implemented the necessary controls to allow European DSs to exercise their "Right of Access" through data requests (DRs), as this right has been extended from the original Directive 95/46/EC originating from 1995. However, the modi operandi and efficacy of these controls in context of information security and privacy has, to the best of our knowledge, not been investigated in current literature. In this paper, we address exactly this issue. More concretely, we examine the following aspects of the "Right of Access":

- Which information about the DS is requested by the DC in order to verify their personal identity?
- Based on the provided information, how does the DC verify the credentials and hence the authenticity of the request?
- Can the requested information be forged by an adversary or can the DC be persuaded through social engineering such that unauthorized access to the DS's personal data is obtained?
- How can the verification of the personal identity of the DS be improved?


*USENIX Symposium on Usable Privacy and Security (SOUPS)* 2019.
August 11–13, 2019, Santa Clara, CA, USA.

The structure of the paper is as follows. In Section 2, we discuss the general format of a DR and how it can be used to exercise the "Right of Access". Section 3 then presents an experiment where we submitted forged DRs to 55 organizations in order to answer the research questions outlined above. Next, we propose a number of possible policy improvements for handling DRs that could be implemented by organizations in Section 4. Moreover in Sections 5, 6 and 7 we respectively discuss related work, limitations and future work, and the conclusions of this study. Finally, a more detailed discussion of the individual cases of our experiment is provided in Appendix A.1.

## 2 The GDPR Data Request

The "Right of Access" [2, Art. 15] introduced by the GDPR allows European consumers to request personal information from any organization that processes their data [1]. As stated in [2, Art. 4-1], "personal data" means any information relating to an identified or identifiable natural person. Practical examples of such personal data can exist of, for instance: location history, financial transactions, written messages, etc.

To exercise this right, the DS has to submit a DR to the desired organization by any means, such as email or postal mail [2, Art. 12]. As the DC should avoid leaking personal data to unauthorized adversaries, it can respond to a DR by requesting the subject to verify their identity and thus ensure that the sensitive data is delivered to the right person.

Each DC should respond to a DR with the requested information, without undue delay and in any event within one calendar month, unless an additional extension of 2 months is requested by the DC due to the complexity or the large number of current DRs [2, Art. 12.3]. This means that the subject should, in any event, at least receive a response within one calendar month and should receive the required information in no more than 3 calendar months, preferably in an electronic format [2, Rec. 59]. Furthermore, the personal data should be presented to the subject in a "commonly used electronic form" [2, Art. 15-3] and in some specific cases, also in a "structured, commonly used and machine-readable format" [2, Art. 20], meaning that – for instance – screenshots are not allowed.

In order to manage such rights effectively, a Data Protection Officer (DPO) should be appointed in organizations whose core activities consist of regular and systemic monitoring of DSs on a large scale or consist of large scale processing of sensitive data [2, Art. 37].

## 3 Data Request experiment

In this section, we discuss an experiment where we attempt to send unauthorized DRs by impersonating targeted individuals and therefore abuse the GDPR "Right of Access". First, we describe the assumptions from our adversarial model and lay out the communication and relations between the authors and targeted individuals in Section 3.1. Moreover, the methodology and ethical aspects on how our experiment was conducted are discussed in Section 3.2 and Section 3.3. Furthermore in Section 3.4, we analyze the different credentials that organizations request in order to verify the identity of the DS. Finally in Section 3.5, impersonation techniques are presented that can be applied to extract or forge credentials from the targeted individuals in practical scenarios.

### 3.1 Adversarial model

We acquired the permission to set up the experiment with 2 of our co-authors (which we will refer to as 'targeted individuals'). Our goal is to impersonate these individuals in order to obtain personal information by performing illicit DRs. First, in order to familiarize ourselves with the targeted individuals, we asked each one of them the following questions:

- The name of the targeted individual.
- A list of several (local, national or international) organizations of which they knew the organizations had personal information regarding them.
- A link to one public social media profile of the targeted individual.
- The home and email address of the targeted individual.

As we will discuss in Section 3.5.1, such information can be easily gathered from various public sources such as social media or government registers. For our two targeted individuals, we indeed found all information listed above on public sources, except for the home address. In practice, an adversarial model may be weakened or fortified depending on the relation between adversary and targeted individual.

From our targeted individuals, we collected the names of 55 unique organizations to which we posed DR as part of our study. Among these organizations, almost half of them are also present in the Belgian Alexa top 50 [3].

As described above, each of the targeted individuals has also cooperated in the composition of this study as an author of this paper. The reason for this is twofold: (1) due to our willingness to perform an ethical experiment, we were uncertain of the scope of personal data that we would receive from external volunteers when performing illicit DRs, hence minimizing an impact on privacy; (2) in a recent framework such as the GDPR, it would be useful to first analyze how different organizations handle DRs. As the DR procedure should not differ significantly between DSs, we focus on the sample size in terms of the number of organizations instead of the number of individuals we targeted.

[1] The GDPR is also applicable for EU organizations that process personal data from non-EU consumers.

## 3.2 Evaluation methodology

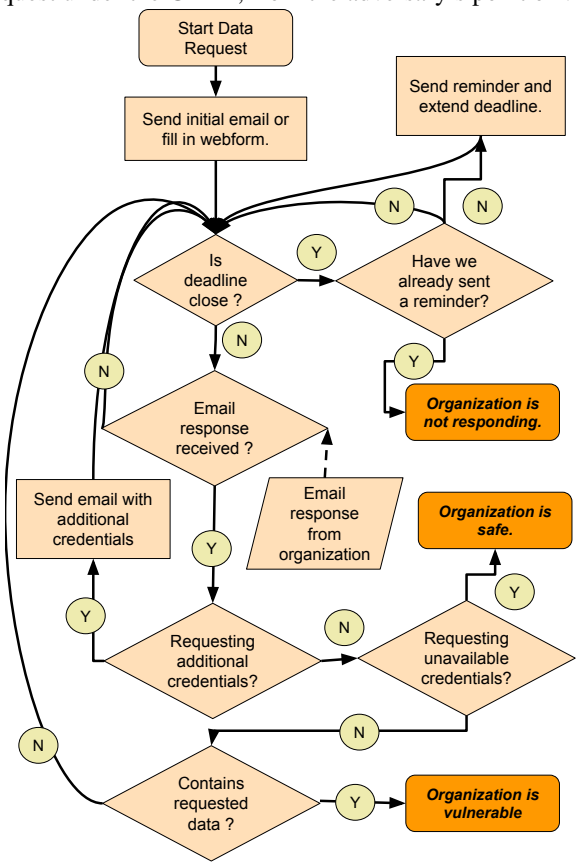
With the list of organizations from each targeted individual, we located the websites of each organization and manually extracted an email address (often located in privacy policies) or link to a web form that is provided to submit DRs. After the extraction, we created a template to exercise the "Right of Access" under the GDPR and submitted a DR to each of the organizations, either through email or by filling in the web form (which we will discuss in Section 3.5.2). With the intention to construct a credible DR, our template also included several questions regarding the retention period of personal data, automated profiling and various methods on how they collect personal information. In the remainder of the paper, the authors are henceforth represented as the adversary, while the targeted individuals are portrayed as the DS.

Our process of performing Data Requests is demonstrated in detail in Figure 1. All email communication was conducted starting from October 16th 2018 until March 12th 2019. Emails that were received on the original email address of each individual, inaccessible by the adversary, were ignored. At the end of the experiment, each organization is assigned to one of the following 3 groups:

- **"Organization is not responding"**: If the organization refrains from responding to our request after a reminder and 2 months of silence, we conclude that the organization is unwilling to fulfill our request and thus is legally not compliant to the GDPR, risking corrective actions (such as fines) [2, Art. 83] and judicial proceedings [2, Art. 79] .
- **"Organization is vulnerable"**: If the organization has delivered personal data from the targeted individual to the adversary, we then conclude that the organization is not able to correctly verify the identity of the DS. As a result, this leads to a data breach of personal information and is therefore non-compliant with the GDPR [2, Art. 88]. Consumers that utilize the services of those organizations are clearly exposed to leakages of their personal information to any determined adversary.
- **"Organization is safe"**: Organizations that do not release personal information about the targeted individual due to secure authentication mechanisms, are considered safe in the context of our adversarial model.

There are 2 exceptions to Figure 1, (1) if an organization adheres to the DR by responding to the original email address instead of the email address of the adversary, we consider this organization to be safe as long as the subject's data is not received by the adversary; (2) if the credentials requested by the organization are not available to the adversary (indicated by "Requesting unavailable credentials"), then we attempt to persuade the DC using the techniques presented in Section 3.5.4. Furthermore, deadlines of one month are

Figure 1: Our experimental process of performing a Data Request under the GDPR, from the adversary's point of view.




established unless the organization requests to extend the deadline with two months, corresponding to Article 12 [2, Art. 12]. Moreover, in case the company is considered to be safe, we assist the targeted individuals to continue the DR process in order to analyze the personal data for any incidental leaks.

On the grounds of ethical research, we do not publicly denounce organizations by name and therefore use a pseudonym that indicates the category in which the organization belongs. These categories consist of: Financial (Fin_x), Retail (Ret_x), Entertainment (Ent_x), Transport and Logistics (Trl_x), News Outlet (New_x) and Other (Oth_x) organizations.

## 3.3 Notes on ethical research

In compliance with ethical research guidelines, the experiment performed in this study was approved and authorized by the university Ethical Research Committee (ERC). Involved individuals were required to acknowledge, through a signed declaration, that their credentials would be used in order to submit unauthorized DRs. Moreover, the targeted individuals (co-authors) gave written permission to read any relevant email communications between them and the DCs for the duration of the experiment. Furthermore, the personal data that

we unintentionally received from the organizations regarding unrelated individuals, were immediately removed after taking note of the event. In addition, a copy of the data from the targeted individuals was sent to the rightful individuals and removed by the adversary after the experiment was finished.

Similar to a responsible disclosure model [8], all "vulnerable" organizations have been notified of the details concerning our research and were individually given advice via email on how to improve their policies of handling DRs. This interaction led to a follow-up personal meeting with the Data Protection Officers of three organizations, where the findings and suggestions for improvements were discussed more in-depth. Our approach to this study was appreciated by the DPOs, as we further ensured that the vulnerable organizations had a reasonable amount of time to implement any necessary changes to their process before publication of this study.

As we will discuss in Section 3.4.3, part of our experiment involved modifying an individual's proof of identity before sending it to an organization. It should be pointed out that no official government documents were altered during this process, only a scanned photocopy. At the same time, we acquired prior permission of the individuals whose proof of identity was used and explicitly obtained clearance from our legal council and the ERC.

Furthermore, we recognize the fact that processing a DR may lead to a certain financial cost for those organizations that handle them manually or have a significant amount of personal data about the DS. The DRs we sent out in this experiment could be considered needless and thus obtrusive to the organizations involved. To counterbalance this, we opted to contact the organizations afterwards to inform them about the outcome of the experiment and to inform them on potential improvements in their handling of DRs. This way of working was universally appreciated by all organizations involved. At the same time, it should be considered that the only way to obtain the necessary information about practical handling of DRs is by actually sending them out - these experiments cannot be performed in a confined lab context. The authors feel that the societal benefits of improving consumer privacy and the organizations' internal policy (which hopefully will be the long term outcome of this study) outweigh the financial costs.

We strongly recommend that future studies should take these ethical considerations into account when deploying such experiment on a larger scale. Finally, the considered organizations were *not* reported to third parties (e.g. the Data Protection Authority), and their identity was anonymized in this paper in order to minimize reputational damage and the risk of criminal targeting.

## 3.4 Authentication credentials

When a DR is submitted to a DC, the identity of the DS must be verified in order to prevent leakage of personal information to an unauthorized third party. The GDPR therefore suggests that the same authentication mechanism should be used for both DRs and for authenticating the DS to the online services offered by the DC [2, Rec. 57]. However, this practice is not explicitly enforced by law.

Recital 64 additionally states that "the controller should use all reasonable measures to verify the identity of a DS who requests access". Hence, organizations are given the freedom to choose their own policies, depending on their definition of "reasonable measures". This is corroborated further by Article 12, which states: "where the controller has reasonable doubts concerning the identity of the natural person making the request [...], the controller may request the provision of additional information necessary to confirm the identity of the DS." [2, Art. 12 (6)].

In summary, although the GDPR provides general guidelines, the precise type of information that should or should not be requested from the DS for authentication purposes is left to the discretion of the DC.[2] Over the course of the experiment we observed that in practice, organizations indeed request a wide variety of credentials to confirm the identity of their users as a result (the nature of which is typically found in the privacy statement).

In Table 3, we present an overview of all the manually contacted organizations and which credentials (authentication data) they requested in order to verify the identity of the DS. Additional details related to this table are defined as followed:

- The "Link leakage" check marks indicate whether the organization unintentionally leaked other personal information unrelated to our initial DR, which can occur in 2 cases: (1) personal data from other individuals with a similar or identical name are included in the response to a DR; (2) the organization has no email address for some account A on file, so the first account B that is created with a name and date of birth identical to account A, will be linked by the organization to account B. An adversary is able to create account B and then perform a DR for account B, resulting into a leakage of data from account A through account B.

- A check mark in the "Vulnerable" column corresponds to the "Organization is vulnerable" description, as discussed in Section 3.2.

- The column "Region" indicates the organization's market area, defined to be either "Local", "National" or "International".

The following subsections discuss the results of this table and describe the different types of required credentials that we encountered in detail.

[2] Subject to the general principles of processing personal data contained in article 5 of the GDPR, such as data minimisation.

Table 1: Number of automatic and manual DRs handling processes of organizations, including the number of answered and unanswered DRs and the number of vulnerable organizations.

| DR process | Answered | Unanswered | Vulnerable |
|---|---|---|---|
| Automatic | 14 | N/A | 0 |
| Manual | 37 | 4 | 15 |

#### 3.4.1 Login credentials

In Table 1, we show that for 14 out of 55 investigated organizations, performing a Data Request is *only* possible through a dedicated web page after logging in on the organization website, as recommended in Recitals 57 and 63 of the GDPR [2, Rec. 57, 63]. An extra 3 out of the remaining 41 organizations require the DS to log in (e.g. through an external dedicated webpage of privacy management software) after the identity was verified through email communication, which is referred to as "Account verification" in Table 3.[3] In addition, one organization allowed the DS to access their personal data in multiple ways, including login credentials and another organization was persuaded to provide an alternative for the "Account verification" (shown by '*'). In summary with this type of login credentials, the DC provides only the personal data from the account associated with the credentials in question.

Observe from column "Account verification" in Table 3, that all DCs which require the user to log in are *not* vulnerable, since in these cases the DR procedure is protected by the authentication mechanism of the website. Clearly, these requirements cannot be enforced if the organization does not have a website or if data about the DS was stored without requiring the creation of an account on the organization's website. Either of these scenarios give rise to a significantly greater challenge to verify the identity of the DS, as we will discuss in Section 4.

#### 3.4.2 Email address

Instead of requiring the user to log in, 41 organizations allow the subject to perform a Data Request by explicitly emailing the DPO or DC, whose email address is typically found in the privacy statement on the organization website. As such, the request is manually handled or at least, analyzed by a human correspondent. The DPO/DC should ideally only adhere to the request if it is made from the same email address with which the user is registered on the organization's website.

However, only 12 of 41 organizations enforced this policy and an additional 5 organizations permitted the subject to offer other credentials if the subject no longer has access to their original email account. In most cases, specific user data (e.g. last products bought) was requested to compensate for not being able to access the original email account. Considering that a realistic adversary has no access to such information, and assuming this information cannot be trivially guessed, we consider these organizations to be safe unless a link leakage has occurred such as for example in the case discussed in Appendix A.1.3.

[3] This is different compared to the "automatic" process, as such a process does not allow a DS to initially request their personal data by email.

#### 3.4.3 National identity card

Another credential required by 13 out of 55 organizations is a digital copy or scan of the national identity (ID) card of the subject. The copy is either uploaded via a web form dedicated to DRs or included as attachment in case the DR is performed via email. One organization requested the front and back side of the ID card, while the remaining 12 organizations only requested the front side. Note that while the National Register Number (NRN) is only written on the back side, the Card Identification Number (CIN) is located on the front side of the ID card. However, since "a controller should not retain personal data for the sole purpose of being able to react to potential requests" [2, Rec. 64], sensitive data on the ID card that is known not to be in possession of the DC, e.g. NRN and CIN, can be censored by the subject [2, Art. 25]. In fact, this was explicitly required by default for 11 organizations.

#### 3.4.4 Home address

A lesser used credential is the home address of the subject, required by 5 out of 55 organizations. Four of these request the complete address consisting of the street name and city, while the remaining organization only demands the region in which the subject lives such as the city or province. Generally speaking, knowing the region of the subject is a relatively easy task for a determined adversary given that social media accounts often disclose this information; it can also be obtained through various public databases as we will discuss in Section 3.5.1. Likewise, even the complete address of the subject might be available (although this information is typically contained in other sources).

Forms of Human Intelligence (HUMINT), where the adversary might be able to communicate directly with the subject or friends of the subject, is also a valuable approach to steal the necessary information. Phishing campaigns are clearly an effective method to extract such personal information.

#### 3.4.5 Calling the subject

Calling the subject on a phone number known by the DC beforehand is a safer authentication method, but is unfortunately only carried out by 2 out of 55 organizations. By making a call, the DC can speak directly to the DS and as such confirm the submission of a DR or request additional user-specific data for the purpose of authentication (see for example Appendix A.1.2).

For an adversary, intercepting calls to the DS's phone is difficult, although possible through for example additional so-

cial engineering [6]. On the other hand, spoofing the caller ID of the subject is a relatively trivial task [11], but has no useful purpose in this scenario as the DC calls the subject and not the other way around. In case the subject performs the initial DR orally (for instance; through a phone call), the DC must still verify the identity through other means [2, Art. 12.1], presumably to avoid precisely such an identity spoofing attack.

As we had no access to the mobile phone of the targeted individuals, we concluded that organizations that performed this authentication method are safe in the context of our adversarial model.

#### 3.4.6 Specific user data

The final credential that we discuss is a demand of the DC to provide specific user data from the DS, requested by 11 out of 55 organizations. This includes various unrelated pieces of information, depending on the nature of the organization. For instance, an entertainment venue might ask to provide the date of the last visit and the products that were bought by the DS.

Determining this information for an adversary is challenging and usually requires in-depth knowledge of the DS. Here, Open Source Intelligence (OSINT) methodologies are useful to e.g. find photos that indicate visits but are in many cases not sufficient to discover the exact details required. Due to difficulty of extracting the necessary information, we consider organizations that request such specific data to be safe in the context of our adversarial model.

### 3.5 Impersonation techniques

In order for an adversary to obtain the subject's personal data through a DR, they must trick the DC into believing the request is legitimate by impersonating the subject. Due to the non-explicit nature of Recital 64, there is no one-size-fits-all approach to achieve this goal (which is also true for social engineering in general), and a determined adversary is more likely to devise an impersonation strategy that is specifically tailored to meet the set of requirements mandated by one specific organization. In this section, we discuss the impersonation techniques useful in forging or extracting the necessary credentials.

#### 3.5.1 Intelligence gathering

As impersonation strategies often demand information from external sources, we explore a number of different intelligence techniques that are able to fabricate a trustworthy profile of our targeted individual. In this section, we merely explain the possible methods of extracting basic information useful to perform illicit DRs.

The most common approach is Open Source Intelligence (OSINT), a form of collecting publicly available information from the targeted individual. Especially in society today, social media plays an important role in extracting personal data. Unsurprisingly, 79% of all people that have Internet access are in possession of at least one social media account [16]. Various social media platforms have different pieces of sensitive information depending on how strong an individual has chosen to shield that information.

For instance, a basic public version of a social media profile often consists of numerous personal images that could be used to alter a photo of an identity card. In cases where the adversary is able to open up the profile by either requesting to become friends or following the targeted individual, sensitive information becomes much more accessible. For example, the date of birth or the region of residence often becomes visible, which is information essential to employ impersonation strategies. In some extreme cases, images of purchase deeds or result sheets of driving examinations are uploaded which clearly display the address of the targeted individual. Additional leakages are also possible by discovering matches between what people like or analyzing the social media profiles of relatives [22]. Besides social media platforms, central government agencies such as the NBB (National Bank of Belgium) or telephone directories such as De Witte Gids also contain personal information (often publicly accessible).[4] Another possibility is to utilize global OSINT search engines such as Pipl [25], which permit adversaries to collect a significant amount of data from individuals with minimal effort.

As opposed to OSINT, a more rigorous and tedious approach called HUMINT is also viable to extract sensitive information from a targeted user. HUMINT serves as the basis for phishing campaigns, in which unsuspecting victims are contacted and then tricked into releasing personal identifiable data by using social engineering techniques [21]. In the context of our proposed impersonation strategies, only weak phishing campaigns are necessary where the targeted individual is able to provide us the personal information we require. However, not only the Internet is a profitable source for personal information; television and public appearances may also increase the risks of extracting valuable intelligence related to e.g. public figures.

Another source of information available to the adversary could stem from a possible personal relation with the targeted individual. For instance, a spouse may already have a significant amount of information available and therefore, would not be required to perform any lookups on social media. In fact, close relatives that reside in the same household such as a spouse, brother or sister might even be able to access the smartphone of the individual, thereby circumventing the "Call subject" authentication method. To the contrary, a person unknown to the targeted individual may not have access to the physical address and therefore has to consult additional sources to collect this information.

To conclude, we argue that excerpting enough personal

[4] "De Witte Gids" (https://dewittegids.be) is the Belgian version of a "White Pages" directory.

identifiable information from a socially active user is feasible, given the many possibilities for a determined adversary.

### 3.5.2 Email address spoofing

A common and basic strategy to impersonate a user (subject) is to spoof their registered email address, which we will henceforth refer to as the *original email address*. Any email address controlled by the adversary will be denoted as a *fake email address*. In our experiment, we applied a number of techniques to impersonate the targeted individuals via email:

- **"The Reply-To"**: The adversary sets the `From` header of the email to the original email address and the `Reply-To` header to a fake email address before sending. Upon replying to this email, email clients should automatically fill in the email address from the `Reply-To` header as the destination.[5] Furthermore, at the time of writing, most popular email clients (for example Gmail and Outlook), only show the `From`, `To`, and `CC` fields to the user when an email is opened, whereas the `Reply-To` field is hidden by default. As a result, an inattentive handler of the DR could be tricked into thinking that the DR originated from a legitimate user, while the reply is sent towards an email address under control of the adversary.
- **"The Resembler"**: The adversary registers a domain name that is similar to the domain of the original email address by using homographs. This is similar to the homograph attack described in the work of Gabrilovich and Gontmakher [12], except that the letters need to be in the same script as per ICANN guidelines [18, p. 2]. The DR is then sent from a fake email address on this domain.
- **"The Ringer"**: The adversary creates a fake email address that is identical to the original email address except for the domain, and sends the DR using this email. For example, if the original email address is "john.doe@gmail.com", the adversary will send the DR using "john.doe@protonmail.com".

Although in our experiment we only employ these techniques exactly as described above, it should be noted that in practice, many variations could be improvised. As an example, consider the case where an adversary uses "The Resembler" technique to submit a DR through a spoofed email, except this time they do not register the homographic domain name. This will render the organization unable to respond to the DR, as the domain name is not registered. Next, after a certain period of time (for example 30 days), the adversary sends a reminder email from a different email address under their control, which cites the first DR request that was transmitted with the spoofed email address. Upon receiving the reminder, the DC may recall that they were indeed unable to reply to the first DR, and be inclined to respond to the reminder email. This is exacerbated by the fact that the citation of the original email may give a false sense of legitimacy to the reminder email, despite that it was sent from a different email address under control of the adversary.

[5] It should be noted that RFC 2822 does not explicitly require that replies must be sent to the `Reply-To` address [26].

Table 2: Brief experiment to choose the best impersonation strategy by sending a DR to 15 organizations (5 per technique) and count the received responses to the adversary email address.

| Technique | Received | Not received |
|---|---|---|
| The Reply-To | 1 | 4 |
| The Resembler | 4 | 1 |
| The Ringer | 5 | 0 |

Continuing our study, the question now remains which impersonation strategy should be chosen by the adversary and how much information should be included in the original DR in order to maximize the probability of success. For finding the best email spoofing technique from the techniques discussed above, we performed a brief experiment involving 15 organizations, where each of the email spoofing techniques was used to contact 5 organizations. The results of this experiment are shown in Table 2.

As shown in the table, the "Ringer" technique resulted in the highest probability of receiving a reply to the adversary's email address, whereas the other techniques were less successful. This may be attributed to a number of reasons: the "Reply-To" technique fails if the DR email is forwarded to another person, in which case the Reply-To header is dropped. Similarly, the header may be dropped if the organization uses a ticketing system for handling emails. In these cases, the replies to the DRs were sent to the original email address instead of the fake one. Another disadvantage of the "Reply-To" technique is that it cannot be used if the organization uses a web form to submit DRs.

For testing the "Resembler" technique, one could attempt to register the domain name `protonmail.com` (Cyrillic `a`), which is similar to the domain `protonmail.com` of an account owned by one of the targeted individuals. However, this approach would fail because registering mixed-script domain names is disallowed by ICANN [18, p. 2] for the purpose of countering homograph attacks. Instead, we registered the domain name `protonmaíl.com` (note the accented 'i'), which contains letters that all belong to Latin script. Similarly to the "Reply-To" case, we noticed that some replies to the Data Requests were sent to the targeted individual's email address. If the email is handled manually by a customer service representative, this may occur if the reply's destination email address is typed manually or if it is corrected by the representative. Moreover, if the organization uses a web form, the DR was in some cases rejected altogether because of the invalid character í in the domain name.

Figure 2: A "John Doe" example of our altered ID card. Metadata of the PNG file such as the image dimension was also modified to increase the credibility of it being captured by a real photo camera.


The email spoofing types used for each organization are abbreviated in Table 3 as "Res", "Rin" and "Rep" for respectively The Resembler, The Ringer and The Reply-To.

#### 3.5.3 Identity card image manipulation

Recall from Section 3.4.3 that if a photo or scan of the front side of an ID card is requested as an authentication credential, sensitive information such as the CIN, hand written signature and validity date number can be censored. Any information that could technically be used as a unique authentication credential that is unknown to the adversary is thereby removed. Consequently, for an adversary to successfully alter a digital copy of an ID card, only the subject's name, photo, and birth date must be known. Though physical ID cards are designed to be difficult to fabricate, a digital copy can be trivially falsified using image manipulation software as depicted in Figure 2.

In this experiment, we replaced the name, birth date, and photo on a reference ID card image to the credentials of the targeted individuals so as to create a manipulated ID card image. The targeted individuals' credentials (photo and date of birth included) were obtained through OSINT from one of their social media accounts. Using the altered ID card image, we were able to successfully authenticate as the targeted individuals in 7 out of 13 organizations that requested the ID card as part of the DR. The remaining 6 organizations requested, in combination with an ID card, additional credentials which we were not able to forge.

Despite having used a legitimate photo on the ID card of each targeted individual, it is unclear whether organizations that have a photo of the subject on file, actually compared them. If not, a stock photo could have been used, thereby reducing the number of known credentials even further and simplifying the process of creating an altered ID.

Clearly, a digital image of an ID card is not an optimal credential for authenticating users. Ignoring the privacy risks, even if the DC would ask for an uncensored NRN and be able to verify it, using the ID card as a credential would still be insecure: if leaked, NRN cannot be changed. Unfortunately, such events have occurred in the past before [5, 9].

#### 3.5.4 Social engineering

Even when a DC requests credentials that are unavailable to the adversary in order to verify the identity of the DS, we found that in practice, the handler of the DR can sometimes be persuaded to offer alternative verification methods through social engineering. The success rate of this approach depends on various factors, including: the personality and current mood of the DR handler [28], the flexibility of policies implemented by the organization and whether employees are trained to recognize social engineering attempts [15].

Specifically in our study, we were able to persuade 8 out of the 41 DCs that handle DRs manually to diverge from standard procedure. In general, we employed the following strategies to attempt to persuade the DC:

**Dismissing access to the DS's email address**: When the DC requires that the DS's registered email address must be used to request or to receive personal data, the adversary can attempt to avoid this requirement by stating that they "no longer have access to this email address". For Ent_A, Ent_D, and New_B, an alternative verification method was offered where the adversary was asked to provide specific user data (which they do not have). These organizations are therefore not vulnerable. Ret_B allowed the adversary to provide an ID card as an alternative, but always required the user to log in to actually download their personal data. Trl_C on the other hand sent the DS's data to the adversary's email address without any additional verification.

**Dismissing access to the DS's online account**: If the DC sends the requested data via the online platform of the organization, as is the case for Fin_C, the adversary cannot retrieve the requested data. In such a scenario, the adversary can pretend that the requested data was never delivered by sending a reminder email. Fin_C responded to this by sending the requested data again via postal mail to the DS. Although the adversary also cannot intercept the DS's postal mail, the established trust with the DC allowed the adversary to request for the rectification of personal data (see Appendix A.1.1 for details of this interaction). Interestingly, Fin_C's online platform implements a two-factor authentication mechanism for logging in, and as such the adversary essentially managed to bypass this mechanism by performing a DR.

**Deliberately omitting unknown credentials**: The DC Fin_A by default requires the DS to provide both the front and back side of their ID card. Because the back side of a Belgian ID card contains the NRN, which the adversary does not know, the adversary requested to omit this information "due to pri-

vacy concerns". More specifically, Ent_L required a product serial number, name and date of birth of the DS, of which the adversary simply omitted the product serial number without further explanation. Since the GDPR is not explicit in stating which credentials are sufficient [2, Rec. 64], we postulate that Fin_A and Ent_L agreed to provide the personal data to the adversary in light of maintaining positive customer relations.

Naturally, social engineering is only possible if the organization allows the adversary to interact with a person at some point during the DR handling process. Therefore, the risk of successful persuasion through social engineering can be mitigated by implementing an automated DR handling process that can be initiated by the DS upon successfully authenticating on the organization website, as described in Section 3.4.1.

## 3.6 Types of personal data leakage

In the previous sections, we outlined the various credentials requested by organizations in order to verify the identity of the DS and demonstrated how an adversary can use impersonation techniques to pass the verification process in the interest of obtaining personal data of a targeted individual. We will now present an overview of the various types of personal information that were leaked by the organizations considered in our study. Since listing the types of data leaked by each organization individually could reveal the identity of the organization, we group the personal information leakages per organization category:

- **Financial institutions**: ID card number, list of timestamped financial transactions, customer ID, telephone numbers, place of birth, partial debit and credit card numbers, list of products purchased from the financial institution, and account numbers.
- **Retail**: List of purchased products, information on purchased products (e.g. serial number), sold products, and delivery dates.
- **Entertainment**: Purchased products and preferences.
- **Transport and logistics**: Timestamped visited locations with GPS coordinates, saved routes, purchased tickets, purchased subscriptions, and customer ID.
- **News outlets**: Browsing history, personal preferences and information about the device used to visit the news outlet's website (e.g. browser and operating system).

Aside from the personal data listed above, each category also leaked the full name, home address and email address of the targeted individual. The data was delivered to the adversary via email as either a `pdf`, `csv`, `xls`, `doc`, text or screenshot attachment. Note that the information obtained from each of these organizations could in practice be "daisy chained" to increase the credibility of DRs to other organizations, although we did not consider this type of adversary in our study.

## 3.7 Summary of results

As shown in Table 1, we have analyzed the policies of 55 organizations, of which 14 have an automated process and 41 process requests manually. From the latter, there were 4 organizations that did not respond to our DR, even after repeated attempts.

None of the organizations with an automatic process had any "Link leakage". However, 15 out of the 41 manually contacted organizations have leaked personal data from the targeted individual to an unauthorized third party. Ignoring the organizations with a "Link leakage", there are still 12 organizations that are left vulnerable to illicit DRs.

Interestingly, financial organizations – which should have a higher responsibility and higher standard of compliance required to safeguard personal information – are vulnerable in 4 out of 5 considered organizations, as shown in Table 3. To the contrary, only 2 out of 12 considered entertainment organizations are vulnerable. From the total of 15 vulnerable organizations, there are 8 organizations which would not have been vulnerable without an altered ID card. Meanwhile, the remaining 7 organizations were exploitable by persuading the DR handler or by using extracted information from OSINT sources.

# 4 Improving Data Request authentication

Based on the findings of our study, we propose several recommendations for organizations on how to securely handle a DR and for consumers on how to protect themselves against identity theft in the context of DRs.

## 4.1 Recommendations for organizations

Our results have shown that a substantial number of existing GDPR policies that implement authentication methods for DRs are clearly inadequate. Nevertheless, Recital 57 of the GDPR [2, Rec. 57] suggests DCs to verify the identity of the subject by offering a dedicated service where a subject is able to authenticate him/herself by providing the same credentials used for the online platform of the DC. From a technical viewpoint, we agree that the suggestion in the current recital is an effective method, as there would be no increase in risk resulting from having a separate authentication mechanism specifically for handling DRs. Due to the automated nature of such a service, it also minimizes the risk of link leakages and social engineering.

Despite this being a useful method, small to medium-scale organizations usually do not have the resources to realize such a service as it often requires expensive architectural changes in order to build them in a secure and reliable way [23]. In case an organization is unable to build the aforementioned service but still has knowledge of an email address of the subject, we suggest the DC to strictly adhere to a policy of accepting DRs only from precisely this registered email

Table 3: Overview of requested credentials and the resulting susceptibility for leakages from all 37 organizations that responded to our manual DR. The columns denote the required credential, while the rows indicate the pseudonyms of each considered organization. An asterisk shows that the corresponding credential was not forced, by either accepting an alternative credential or by being able to persuade the DC (*).

| Organization | Account verification | Access to user email | Date of birth | Region of residence | Address | Front ID | Back ID | Call subject | Specific user data | Link leakage | Vulnerable | Region | Spoofing type |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Fin_A | | | ✓ | | | ✓ | * | | | | ✓ | N | Rin |
| Fin_B | | | ✓ | | | | | | | | ✓ | N | Res |
| Fin_C | * | | ✓ | | | ✓ | | | | | ✓ | I | Rin |
| Fin_D | | | ✓ | | | ✓ | | | | | ✓ | I | Rin |
| Fin_E | | ✓ | ✓ | | | | | ✓ | ✓ | | | I | Rin |
| Ret_A | | | ✓ | | | | | | | | ✓ | L | Rin |
| Ret_B | ✓ | * | ✓ | | | * | | | | | | I | Rin |
| Ret_C | | ✓ | ✓ | | | | | | | | | I | Rin |
| Ret_D | | | ✓ | | | | | | | | ✓ | N | Res |
| Ret_E | | | | ✓ | | | | | | | ✓ | N | Res |
| Ret_F | | | ✓ | ✓ | ✓ | ✓ | | | | | ✓ | I | Rin |
| Ret_G | | | | | | | | | ✓ | | | N | Rep |
| Ret_H | | ✓ | ✓ | | | | | | | | | I | Rin |
| Ret_I | | | ✓ | | | | | | | ✓ | ✓ | N | Rin |
| Ent_A | | * | | | | | | | * | | | I | Rin |
| Ent_B | | ✓ | ✓ | ✓ | ✓ | ✓ | | | | | | N | Rin |
| Ent_C | | | | | | | | | ✓ | | | I | Rin |
| Ent_D | | * | ✓ | | | | | | * | | | I | Rin |
| Ent_E | | ✓ | | | | | | | | | | N | Rep |
| Ent_F | | | ✓ | | | | | | ✓ | | | I | Rin |
| Ent_G | | | ✓ | | | | | | | | ✓ | L | Rep |
| Ent_H | | ✓ | ✓ | | | | | | ✓ | | | I | Rin |
| Ent_I | ✓ | ✓ | ✓ | | | ✓ | | | | | | I | Rin |
| Ent_J | | ✓ | ✓ | | | | | | | | | N | Rep |
| Ent_K | | ✓ | ✓ | | | | | | | | | N | Rep |
| Ent_L | | | ✓ | | | | | | * | | ✓ | I | Rin |
| Trl_A | | ✓ | ✓ | ✓ | ✓ | | | | | ✓ | ✓ | N | Res |
| Trl_B | | ✓ | ✓ | | | ✓ | | | | | | N | Res |
| Trl_C | | * | ✓ | | | ✓ | | | | | ✓ | I | Rin |
| Trl_D | | | ✓ | | | ✓ | | | | ✓ | ✓ | N | Rin |
| New_A | | | ✓ | ✓ | ✓ | ✓ | | | | | ✓ | N | Res |
| New_B | | * | ✓ | | | ✓ | | | * | | | N | Rin |
| Oth_A | ✓ | | | | | | | | | | | N | Res |
| Oth_B | | | ✓ | | | ✓ | | ✓ | | | | N | Rin |
| Oth_C | | | | | | | | | ✓ | | | I | Rin |
| Oth_D | | ✓ | ✓ | | | | | | | | | I | Rin |
| Oth_E | | | | | | | | | ✓ | | | N | Rin |

address. Interestingly, some DCs such as the one discussed in Appendix A.1.1 had knowledge of the original email address (as the email address was contained in the data package) but nevertheless did not mandate this policy. However – even if such policy is adhered to – an adversary that has access to the mailbox of the subject, might still be able to bypass any two-factor authentication (which is potentially required when attempting to log in to the service by normal means).

A more concerning issue is the fact that some DCs are not in possession of online credentials (e.g. email addresses and passwords), making it impossible to implement such a policy. In case the organization does however have the phone number of the subject, we propose to call the subject to verify their identity, even though it requires a human operator, which might be an even greater burden on small scale organizations [14].

A final authentication method that we consider is to request user-specific data from the subject. For instance, an electricity company might ask for multiple reference numbers located on one of the subject's invoices, while an insurance company is able to request similar information located on insurance papers. However, care needs to be taken as some user specific data might still be easy to deduce, depending on the type of the organization.

Nonetheless, there are situations where the DC has no useful information to verify the identity of the DS. In these cases, the "Right of Access" does not apply and hence, the DS is unable to perform a DR to that organization, unless the DS provides additional information that enables the DC to verify the identity [2, Art.11]. Moreover, recital [2, Rec. 57] suggests DCs to not retain information that is "for the sole purpose of complying with any provision of this Regulation.". In other words, the DC should not retain personal information from the DS with the *only* purpose to respond to possible DRs, thereby significantly reducing the number of available authentication methods. As a result, it is difficult to propose an authentication method in case the DC has an insufficient amount of information to verify against.

A summary of the authentication methods that we propose are listed below, in decreasing order of importance regarding privacy and viability:

1. An automated process that requires known login credentials, without the ability to bypass an existing 2-factor authentication such as SMS messages.
2. A *strict* policy of only permitting DRs for online accounts that are sent by the email attributed to that and only that account. In addition, call the subject and request specific user data.
3. Call the subject and request specific user data.
4. Request specific user data.

It is evident that proposing a one-size-fits-all approach is problematic. Commercial tools that aid in processing DRs do exist, but it is unclear how reliable and effective those tools are.[6] In conclusion, we also suggest for employees that handle such requests to be trained in how to detect and avoid impersonation strategies in order to securely process DR.

## 4.2 Recommendations for consumers

Regardless of the fact that organizations are primarily responsible for providing improvements, we also present several options for DSs to reduce the chances for future data breaches. As social media is currently used by approximately 45% of all people in the world [16], sharing more personal data poses a significant risk for identity theft in general. Even though removing social media profiles entirely would substantially reduce the risk of illicit DRs, it is often an unrealistic suggestion for many consumers. More realistically, personal information such as profile photos and posts should be hidden from the public and only accessible for (close) friends, thereby shrinking the set of available data to possible adversaries. Many platforms allow consumers to fine-tune their privacy settings separately for each piece of personal information [27]. Nevertheless, we recommend users to completely hide sensitive information that could reveal their date of birth or region of residence on social media platforms as this information may be utilized by adversaries to devise a credible DR.

In addition, consumers should be attentive to emails that contain information related to DRs as they might disclose possible impersonation attempts by adversaries. For instance in Appendix A.1.1, the organization first sends the personal data to the legitimate DS on the online platform, thus indirectly notifying the DS of a failed DR attempt. In this unfortunate event, we recommend consumers to take preliminary measures by contacting the organization in question such that potential data breaches can be mitigated.

As a last recommendation, we suggest consumers to think carefully about the services or products they buy from the corresponding organizations. A quick look at the privacy policy of a given organization might already provide a rough judgement about the importance of privacy in that organization. Furthermore, performing a legitimate DR as a consumer will divulge most of the credentials necessary in the organization's process of verifying the DS's identity. Clearly, requesting credentials such as an ID card or basic personal information may indicate a poor GDPR policy for handling DRs.

# 5 Related work

To the best of our knowledge, Galetta et al. [13] were the first to empirically examine the practicality of performing DRs in Belgium under the now repealed Directive 95/46/ EC [1]. In their work, they showed that DCs were often insufficiently prepared to handle such requests as only 11 out of 19

[6] e.g. OneTrust and Jumio

organizations responded to their initial DR. Two years later, Ausloos et al. [4] confirmed the lack of privacy awareness with another 60 services and further discuss the difficulties that DSs encounter when attempting to exercise their rights.

After the GDPR went into effect, Wong et al. [29] exercised several consumer rights introduced by the GDPR with 230 different organizations and showed improvements in terms of usability compared to previous work, but demonstrated inadequacy in data formats. Furthermore, the authors briefly touched upon the various authentication methods that were required by the DCs in which only 88 out of 230 organizations required additional credentials. Surprisingly, 62 out of 230 DCs did not provide the subject with the personal information that was mandated after a period of 3 months. However, their experiment had a different approach compared to ours as they did not attempt to impersonate other DSs and furthermore, did not discuss the different credentials required in the point of view of an unauthorized adversary.

More recently in 2019, additional studies regarding the "Right of Access" have been conducted to show the negligent behaviour of organizations as some of them still do not correctly adhere to the subjects' rights or flat-out refuse to handle the DRs [7, 10, 24].

Though the "Right of Access" has not been subject to social engineering techniques in related work, there are a number of works that explore such techniques in an OSINT context [17, 20].

## 6 Limitations and future work

Our study has a number of limitations that could be addressed in future work. First, the set of targeted individuals is limited in size, as it consists of two co-authors. Even so, we argue that this limitation does not discredit our findings as an organization's DR handling *process* ideally *should* not differ from one DS to the other. We also postulate that recruiting a large number of participants for similar studies will prove to be a difficult task, since there is a significant risk involved for each participant. Indeed, as mentioned in Section 3.6, a large quantity of highly sensitive information about the participant may leak. The participants must fully trust that such leaks will not be abused by the researchers. If so, considering a larger set of targeted individuals in a future study, with multiple DRs per organization, would reduce the probability of false negatives (organizations that have a poor policy but where the adversary got "unlucky" and information was not leaked). Furthermore, biases towards certain ethnicities, professions or nationalities could be identified. It should however be noted that, with an increased number of DRs directed towards a single organization, additional care must be taken not to raise suspicion.

Second, our study considered 55 organizations, coming from a broad range of industries. Although we believe this ensures the generalizability of our findings, it might be interesting for future studies to focus more on one specific industry in order to discover any characteristic patterns pertaining to that industry.

A final limitation is that our study cannot precisely define the required credentials for a successful DR. This is due to the fact that DR handling processes differ significantly between organizations and are not fully disclosed in a public way. Furthermore, for organizations that make human interaction part of the process, the success of a DR is also dependent on the personality of the DR handler. Subsequent studies may therefore consider the rigorousness of an organization's policies and how those can be transferred and abused in related rights such as the "Right to Rectification" [2, Art. 16].

## 7 Conclusion

In this paper, we have explored the different credentials (authentication methods) that are requested by organizations in order to verify the identity of DSs under the "Right of Access". Additionally, different social engineering techniques have been applied to realistically forge these required credentials.

As a result, we have demonstrated that a significant number of policies for handling GDPR DRs are vulnerable due to either weak authentication mechanisms or the involvement of humans to carry out the processing of the DRs . Out of 55 examined organizations, 15 have leaked sensitive and personal information from the targeted individuals participating in our experiment, including but not limited to financial transactions, website visit histories and timestamped locations. Exercising the "Right of Access" while impersonating a DS is therefore an appealing attack for criminal adversaries.

Furthermore, we have proposed well-established authentication methods to improve the DR policy within the current legal framework. Yet, as some organizations are unable to perform these proposed methods due to not being in possession of the appropriate authentication credentials, we acknowledge that these organizations still run an increased risk of unintentionally leaking personal data to a determined adversary. We conclude that precautions also have to be taken by consumers, as it is possible to obtain valuable information through OSINT, which – as we have shown – might ultimately lead to a substantial impact on privacy.

## Acknowledgements

This research was funded in part by the Bijzonder Onderzoeksfonds (BOF) of Hasselt University and by a Ph.D. Grant of the Research Foundation Flanders (FWO), grant number 1S14916N. Finally, we thank the reviewers and shepherd for their in-depth feedback.

## References


[1] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data). *OJ L 281* (November 1995), 31–50.

[2] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *OJ L 119* (May 2016), 1–88.

[3] ALEXA. Top Sites in Belgium. https://www.alexa.com/topsites/countries/BE, accessed on January 25th 2019.

[4] AUSLOOS, J., AND DEWITTE, P. Shattering one-way mirrors - data subject access rights in practice. *International Data Privacy Law 8*, 1 (03 2018), 4–28.

[5] BARRETT, B. Hack Brief: An Astonishing 773 Million Records Exposed in Monster Breach. https://www.wired.com/story/collection-one-breach-email-accounts-passwords/, accessed on January 27th 2019.

[6] BARRETT, B. How to Protect Yourself Against a SIM Swap Attack. https://www.wired.com/story/sim-swap-attack-defend-phone/, accessed on January 15th 2019.

[7] BONIFACE, C., FOUAD, I., BIELOVA, N., LAURADOUX, C., AND SANTOS, C. Security analysis of subject access request procedures how to authenticate data subjects safely when they request for their data. In *Annual Privacy Forum* (2019).

[8] BUGCROWD. What is Responsible Disclosure? https://www.bugcrowd.com/resource/what-is-responsible-disclosure/, accessed on May 17th 2019.

[9] DAVID VOLODZKO. Marriott Breach Exposes Far More Than Just Data. https://www.forbes.com/sites/davidvolodzko/2018/12/04/marriott-breach-exposes-far-more-than-just-data/, accessed on January 15th 2019.

[10] DEMEYER, S., AND VANRENTERGHEM, A. Wat weten bedrijven echt van u? Het blijft vaak onduidelijk (Dutch). https://www.vrt.be/vrtnws/nl/2019/01/28/privacyonderzoek/, accessed on February 21st 2019.

[11] FEDERAL COMMUNICATIONS COMMISSION. Caller ID Spoofing. https://www.fcc.gov/consumers/guides/spoofing-and-caller-id, accessed on January 25th 2019.

[12] GABRILOVICH, E., AND GONTMAKHER, A. The Homograph Attack. *Commun. ACM 45*, 2 (Feb. 2002), 128–.

[13] GALETTA, A., FONIO, C., AND CERESA, A. Nothing is as it seems. the exercise of access rights in Italy and Belgium: dispelling fallacies in the legal reasoning from the 'law in theory' to the 'law in practice'. *International Data Privacy Law 6* (11 2015), ipv026.

[14] GDPR REPORT. GDPR is being abused by cyber-criminals to breach complacent businesses. https://gdpr.report/news/2018/07/04/gdpr-is-being-abused-by-cyber-criminals-to-breach-complacent-businesses/, accessed on February 10th 2019.

[15] GRAGG, D. A multi-level defense against social engineering. *SANS Reading Room, March 13* (2003).

[16] HOOTSUITE. Global Digital Report 2019. https://hootsuite.com/pages/digital-in-2019.

[17] HUBER, M., KOWALSKI, S., NOHLBERG, M., AND TJOA, S. Towards automating social engineering using social networking sites. In *2009 International Conference on Computational Science and Engineering* (Aug 2009), vol. 3, pp. 117–124.

[18] ICANN. Guidelines for the implementation of internationalized domain names. https://www.icann.org/en/system/files/files/idn-guidelines-02sep11-en.pdf, accessed on January 15th 2019.

[19] INGBER, S. Amazon Customer Receives 1,700 Audio Files Of A Stranger Who Used Alexa . https://www.npr.org/2018/12/20/678631013/amazon-customer-receives-1-700-audio-files-of-a-stranger-who-used-alexa?t=1549965015007, accessed on February 10th 2019.

[20] IRANI, D., BALDUZZI, M., BALZAROTTI, D., KIRDA, E., AND PU, C. Reverse social engineering attacks in online social networks. In *Detection of Intrusions and Malware, and Vulnerability Assessment* (Berlin, Heidelberg, 2011), T. Holz and H. Bos, Eds., Springer Berlin Heidelberg, pp. 55–74.

[21] KROMBHOLZ, K., HOBEL, H., HUBER, M., AND WEIPPL, E. Advanced social engineering attacks. *Journal of Information Security and applications 22* (2015), 113–122.

[22] LAM, I.-F., CHEN, K.-T., AND CHEN, L.-J. Involuntary information leakage in social network services. In *Proceedings of IWSEC 2008* (2008).

[23] LEONID BERSHIDSKY. Europe's Privacy Rules Are Having Unintended Consequences. https://www.bloomberg.com/opinion/articles/2018-11-14/facebook-and-google-aren-t-hurt-by-gdpr-but-smaller-firms-are, accessed on January 25th 2019.

[24] NOYBD. Netflix, spotify & youtube: Eight strategic complaints filed on "right to access". https://noyb.eu/access_streaming, accessed on January 27th 2019.

[25] PIPL. Pipl. https://www.pipl.com/, accessed on January 25th 2019.

[26] RESNICK, P. RFC 2822: Internet Message Format. *Qualcomm Incorporated* (2001).

[27] STAY SAFE ONLINE. Manage your privacy settings. *NCSA* (2019). https://staysafeonline.org/stay-safe-online/managing-your-privacy/manage-privacy-settings/.

[28] UEBELACKER, S., AND QUIEL, S. The social engineering personality framework.

[29] WONG, J., AND HENDERSON, T. How portable is portable?: Exercising the GDPR's right to data portability. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers* (New York, NY, USA, 2018), UbiComp '18, ACM, pp. 911–920.

# A Appendix

## A.1 Discussion of individual cases

As this paper generally only provides a statistical overview of the authentication methods and their subsequent breaches, we would like to present a few examples of email communication between the adversary and DC.[7] In the following sections, we will use *"my"* to describe the possession of the targeted individual while acting as an adversary. Moreover, we indicate the email from the targeted individual as the "original email" and identify the "DS" as the targeted individual.

### A.1.1 International financial institute: Fin_C

The privacy policy of *Fin_C* (DC) states that the front of the subjects' identity card is required to submit a valid DR. As we submitted our request with a successful "Ringer" strategy, all electronic communication (responses included) was established with the fake email address. Moreover, the initial DR contained the name, date of birth and identity card of the targeted individual. An automatic confirmation email was received shortly after. This time-sheet depicts the subsequent communication between the adversary and the DC:

**20/11/2018**: Automatic email confirming the reception of our DR.

**6/12/2018**: The data containing all personal information was received on the online platform of *Fin_C*, which is virtually impossible for an adversary to access as it requires logging into the targeted individuals' account.

**15/12/2018**: In conformity with our process methodology, we sent a persuasive reminder to notify the DC that the legal deadline for responding to a DR is closing in and has a remaining 5 days left.

**17/12/2018**: An email was received from the DC, justifying that the data have already been sent to the aforementioned account. In addition, the DC proposed to provide a "copy of my data".

**17/12/2018**: We responded that we did not receive such data on "my" account and agreed to accept the "copy".

**18/12/2018**: The DC confirmed to send a "copy of my data".

**21/12/2018**: A physical copy from the data was received on the targeted individuals' home address.[8]

**23/12/2018**: Even though, the adversary is not aware of the specific contents of the personal information, they have however the ability to know the type of information that is provided by the DC by issuing a legitimate DR. Therefore, we sent the controller a request (still with the adversary email address) to modify "my" personal information as depicted in [2, Art. 68]. More specifically, we demanded to remove the phone number and modify the education degree.

**24/12/2018**: An email from the DC was received, confirming the modification of "my" personal data.

The DC could not be persuaded to send the personal data to the adversary email. However, a request coming from the adversary email to modify the data was accepted, hence allowing an unauthorized change to the personal data of the subject. As we only exercised our right to modify personal information with *Fin_C*, it is inconclusive to know if more organizations are vulnerable to such attack in this scenario.

[7] Dates are using the little-endian notation

[8] We acknowledge that we did not expect to receive the data by postal mail as we were uncertain about the specific meaning of 'copy of the data'.

### A.1.2 International financial institute: Fin_E

In the privacy statement of *Fin_E*, only an email address to send DRs to was provided, without information about the necessary credentials. Submission of the DR included the name and date of birth of the targeted individual and was carried out with the "Ringer" strategy.

**20/11/2018**: DR was sent by email.

**21/11/2018**: Email was received by the adversary, confirming the reception of the DR.

**04/12/2018**: A response from the DC on the original email was received, containing a summary of answers to the questions asked in the initial DR. In addition, they suggested us to "manually visit" the organization web pages in order to extract the necessary information. Article [2, Art. 68] states that data should be delivered "in a structured, commonly used, machine-readable and interoperable format" and hence, the response clearly violates this article.

**17/12/2018**: As a realistic adversary does not have the knowledge of the previous response received on the targeted individuals' email, we send them a persuasive reminder to indicate the approaching deadline according to [2, Art. 12-3].

**17/12/2018**: The DC responds to the adversary's email, stating that an answer to the DR was already offered and forwarded the email message from 04/12/2018 to us.

**17/12/2018**: Since the adversary is now aware of the original email being sent, we notify the DC of their violation of Article [2, Art. 68] and therefore, request the controller to send "my" personal data in a "machine-readable format".

**18/12/2018**: The targeted individual received a phone call on the number known by the DC. In this phone call, they verified the identity of the targeted individual by requesting the birthplace, original email address and specific account data.

**27/12/2018**: The targeted individual received the personal data (consisting of scanned documents) on the original email.

In this case, we argue that the DC did not receive any DR yet that explicitly mentioned the violation of [2, Art. 68], or the DC has no automatic process in place and attempts to eschew the DR by only providing limited information. Nonetheless in the end, additional verification methods were performed which are very difficult for an adversary to forge as it would require access to the phone number and specific knowledge related to the account of the targeted individual.

### A.1.3 Logistics service: Trl_D

To submit a valid DR, the privacy policy of *Trl_D* states that a copy of the identity card, international passport or driving license is required. Additionally, the DC requested to censor sensitive information such as the photo and NRN. Similar to previous cases, this attack was performed with the "Ringer" strategy and all communication was done through the fake email address of the adversary.

**19/11/2018**: DR with the necessary credentials (as stated in the privacy policy) was submitted through a web form.

**19/11/2018**: Automatic reply was received, providing a ticket number.

**22/11/2018**: A response from the DC was received, asking if the email address of the targeted individual also had to be included into the data package. In other words, the DC indicates that there is an account with a different email address belonging to the targeted individual and therefore, requests if the personal information of this account should also be included.

**22/11/2018**: We replied that the email address is indeed an "old and unused one" and hence request personal information from that account.

**18/12/2018**: All personal data from the targeted individual, including additional data from other individuals with a seemingly similar name was received by the adversary in one large PDF file.

In terms of privacy, there are two breaches: (1) the impersonation strategy succeeded and (2) personal information of three additional users were leaked (link leakage). The first breach occurred rather quickly as the email of 22/11/2018 shows clear signs of the DC already assuming the identity of the DS without performing additional verification. The second breach indicates that including additional sensitive information from 3 other individuals is clearly also a privacy issue, albeit with a different impact compared to the previous cases. This type of breach would even exist if the DS would send a legitimate request, similarly to the publicly known 2018 Amazon Alexa data leak where a DS received voice recordings from an unrelated individual [19].

In our case, the occurrence of such mistake happened most likely due to an inaccurate query. With 2 of the 3 unrelated individuals, the cause was clear as the name of the unrelated individual was exactly the same as the name of the targeted individual. In the remaining case, the unrelated individual whose personal data was leaked had the following data:

- Name: A B
- Address: C-D

while the targeted individual had the following personal data:

- Name: B C
- Address: G

The "address" field of the unrelated individual was erroneously contained in the "name" field. Therefore, the resulting field of the unrelated individual became "A B C-D", thus containing the string "B C", which is precisely the name of the targeted individual.

# An Empirical Analysis of Data Deletion and Opt-Out Choices on 150 Websites

Hana Habib, Yixin Zou[†], Aditi Jannu, Neha Sridhar, Chelse Swoopes,
Alessandro Acquisti, Lorrie Faith Cranor, Norman Sadeh, Florian Schaub[†]
*Carnegie Mellon University & [†]University of Michigan*
*{htq, ajannu, nksridha, cswoopes, acquisti, lorrie, ns1}@andrew.cmu.edu*
*{yixinz, fschaub}@umich.edu*

## Abstract

Many websites offer visitors privacy controls and opt-out choices, either to comply with legal requirements or to address consumer privacy concerns. The way these control mechanisms are implemented can significantly affect individuals' choices and their privacy outcomes. We present an extensive content analysis of a stratified sample of 150 English-language websites, assessing the usability and interaction paths of their data deletion options and opt-outs for email communications and targeted advertising. This heuristic evaluation identified substantial issues that likely make exercising these privacy choices on many websites difficult and confusing for US-based consumers. Even though the majority of analyzed websites offered privacy choices, they were located inconsistently across websites. Furthermore, some privacy choices were rendered unusable by missing or unhelpful information, or by links that did not lead to the stated choice. Based on our findings, we provide insights for addressing usability issues in the end-to-end interaction required to effectively exercise privacy choices and controls.

## 1 Introduction

The dominant approach for dealing with privacy concerns online, especially in the United States, has largely centered around the concepts of notice and consent [56]. Along with transparency, consumer advocates and regulators have asserted the need for consumers to have control over their personal data [22, 28, 41]. This has led some websites to offer privacy choices, such as opt-outs for email communications or targeted ads, and mechanisms for consumers to request removal of their personal data from companies' databases.

Despite the availability of privacy choices, including mechanisms created by industry self-regulatory groups (e.g., the Digital Advertising Alliance [21]) as well as those mandated by legislation, consent mechanisms appear to have failed to provide meaningful privacy protection [15, 57]. For example, many consumers are unaware that privacy choice mechanisms exist [33, 48, 60]. Additionally, past research has identified usability and noncompliance issues with particular types of opt-outs, such as those for email communications and targeted advertising [24, 35, 40, 42, 55]. Our study builds on prior work by contributing a large-scale and systematic review of website privacy choices, providing deeper insight into how websites offer such privacy choices and why current mechanisms might be difficult for consumers to use.

We conducted an in-depth content analysis of opt-outs for email communications and targeted advertising, as well as data deletion choices, available to US consumers. Through a manual review of 150 English-language websites sampled across different levels of popularity, we analyzed the current practices websites use to offer privacy choices, as well as issues that may render some choices unusable. Our empirical content analysis focused on two research questions:

1. What choices related to email communications, targeted advertising, and data deletion do websites offer?
2. How are websites presenting those privacy choices to their visitors?

We found that most websites in our sample offered choices related to email marketing, targeted advertising, and data deletion where applicable: nearly 90% of websites that mentioned using email communications or targeted advertising in their privacy policy provided an opt-out for that practice, and nearly 75% offered a data deletion mechanism. These choices were provided primarily through website privacy policies, but were often also presented in other locations. Furthermore, our heuristic evaluation revealed several reasons why people may find these choices difficult to use and understand. In over 80% of privacy policies analyzed, the policy text omit-


*USENIX Symposium on Usable Privacy and Security (SOUPS)* 2019.
August 11–13, 2019, Santa Clara, CA, USA.

ted important details about a privacy choice, such as whether a targeted advertising opt-out would stop all tracking on a website, or the time frame in which a request for account deletion would be completed. Though a less frequent occurrence, some policies contained opt-out links that direct the user to a page without an opt-out, or referred to non-existent privacy choices. We further observed a lack of uniformity in the section headings used in privacy policies to describe these choices. Compounded, these issues might make privacy choices hard to find and comprehend.

New regulations, such as the European Union's General Data Protection Regulation (GDPR) and California's Consumer Privacy Act (CCPA), aim to address issues with privacy choice mechanisms and include strict requirements for obtaining and maintaining consent for practices like direct marketing, targeted advertising, and disclosure or sale of personal data [25, 50]. Our study contributes a better understanding of the mechanisms websites currently use to provide choices related to these practices, and where they may fall short in helping people take advantage of available choices. Additionally, our analysis provides a foundation for future research into the development of best practices for provisioning privacy choices. These recommendations could build upon changes to the consent experience in the mobile app domain, where research showing the benefits of a uniform interface contributed to changes in permission settings implemented by the Android and iOS platforms [4]. Building new approaches for privacy choice provisioning upon practices that are already prevalent may increase the likelihood of adoption.

## 2 Privacy Choice Regulatory Framework

As background, we provide an overview of current legislation and industry self-regulatory guidelines related to the types of privacy choices evaluated in this study: opt-outs for email and targeted advertising and options for data deletion.

### 2.1 Opt-outs for Email Communications

In the United States, the Controlling the Assault of Non-Solicited Pornography and Marketing (CAN-SPAM) Act of 2003 established national standards for companies that send electronic commercial messages to consumers [29]. It requires companies to provide consumers with a means to opt out of receiving communications, accompanied by a clear and noticeable explanation about how to use the opt-out. Once the commercial message is sent, opt-outs must be available to recipients for at least 30 days, and any opt-out request must be honored within 10 business days. The European Union's General Data Protection Regulation (GDPR) also grants consumers "the right to object" when their personal data is processed for direct marketing purposes (Art. 21) [25]. Furthermore, the California Consumer Privacy Act (CCPA), which will go into effect in 2020, grants California residents the right to opt out of having their personal data sold to third parties, such as for marketing purposes [50].

### 2.2 Opt-outs for Targeted Advertising

Since the early 2000s, industry organizations in the United States and Europe — including the Network Advertising Initiative (NAI), Digital Advertising Alliance (DAA), and Interactive Advertising Bureau Europe (IAB Europe) — have adopted principles and self-regulatory requirements related to practices used in online behavioral advertising [21, 38, 52]. DAA member advertisers are required to provide consumers with the choice to opt out of tracking-based targeted advertising [21]. This requirement applies to data used by the company or transferred to other non-affiliated entities to deliver tailored ads, but not for other collection purposes [46].

The GDPR emphasizes consumers' consent to the processing of their personal data for purposes that go beyond what is required to fulfill a contractual obligation or immediate business interests. In asking for consent, websites should present a clear, affirmative action, and ask visitors for agreement rather than incorporating the consent into default settings, such as pre-checked boxes (Art. 4). Consent should be in an easily accessible form, using simple, clear language and visualization, if needed; if the consumer is a child, the language must be understandable by a child (Art. 12). Moreover, visitors are allowed to withdraw their consent at any time (Art. 7). Nevertheless, the GDPR does not explicitly state that consent is required for targeted advertising, and ambiguity in Art. 6 may provide leeway for companies to claim a "legitimate business interest" and collect data for targeted advertising without obtaining explicit consent [25].

### 2.3 Data Deletion Choices

The GDPR also grants consumers whose data is collected in the European Union the "right to be forgotten." This stipulates that under certain circumstances, companies must comply with consumer requests to erase personal data (Art. 17) [25]. Implementations of the "right to be forgotten" vary from account deletion request forms to the ability of consumers to delete certain information related to their profile.

While no general "right to be forgotten" exists in the United States, some US federal laws contain data deletion requirements for specific contexts. The Children's Online Privacy Protection Act of 1998 (COPPA), for example, requires online services that collect personal information of children under 13 years old to delete it upon parental request [30]. The CCPA will also give California residents the right to request their personal data be deleted, except in certain circumstances, such as when the information is needed to complete an unfinished transaction [12].

## 3 Related Work

Our study builds upon prior work that (1) evaluated privacy control mechanisms; and (2) studied consumer attitudes and behaviors related to data collection and use.

### 3.1 Prior Evaluations of Privacy Choices

The usability of websites' privacy communications and controls has long been problematic [47, 48]. Recent work has shown that privacy policies still exhibit low readability scores [26, 44]. Additionally, most websites fail to provide specific details regarding the entities with which they share data and the purposes for which data is shared [34]. Some consumer advocates argue that current control mechanisms nudge people away from exercising their right to privacy with practices, such as creating a cumbersome route to privacy-friendly options, highlighting the positive outcome of privacy-invasive options, and incentivizing consumers to share more personal data through the framing of control mechanisms [54].

Prior studies have also revealed compliance issues related to privacy control requirements. For example, in the early 2000s the Federal Trade Commissions (FTC) found that privacy controls were not ubiquitously implemented at that time, with only 61% of surveyed websites giving consumers options regarding the collection of their personal information [27]. There is also evidence of noncompliance with the GDPR, as some major websites still deliver targeted ads to European visitors who did not consent to the use of their personal data [19].

However, it seems that companies are adjusting their privacy notice and control mechanisms in response to new legal requirements. Degeling et al. found that, among the more than 6,000 European websites surveyed in 2018, 85% had privacy policies; many websites had updated their privacy policies or started to display cookie consent notices when the GDPR went into effect, likely in response to the GDPR's transparency requirements [20]. Yet, it is unclear whether the changes websites are implementing actually serve to protect consumers. Facebook, for example, was criticized for their post-GDPR privacy changes, as users are still not able to opt out of Facebook's use of behavioral data to personalize their News Feeds or optimize its service [13].

Our analysis primarily focuses on usability issues and does not intend to analyze legal compliance (although the latter is an important direction for future work). Next we highlight key findings of prior usability evaluations regarding email communication opt-outs, targeted advertising opt-outs and data deletion choices, the three types of privacy choices on which our analysis is focused. Our study is the first to survey all three forms of privacy choices in a comprehensive manner through content analysis. Our findings provide an overview of current practices and potential usability pitfalls, with ample implications for making privacy choice mechanisms more uniform and apparent across websites.

#### 3.1.1 Evaluation of Email Communication Opt-outs

Due to the CAN-SPAM Act, many websites offer consumers control over which email messages they receive. An audit of top North American retailers in 2017 by the Online Trust Alliance found that 92% of websites surveyed offered unsubscribe links within messages. However, the study also revealed that compliance issues still exist as some retailers offered broken unsubscribe links, or continued to send emails after the 10-business-days deadline [55]. A 2018 analysis by the Nielsen Norman group revealed usability issues related to unsubscribe options in marketing emails, such as inconspicuous links without visual cues indicating that they are clickable, long and complicated processes involving many check boxes and feedback-related questions prior to the final unsubscribe button, as well as messaging that might annoy or offend users [53]. Our research complements these studies by examining usability issues occurring in unsubscribe mechanisms offered on websites rather than through emails, such as links in privacy policies and account settings.

#### 3.1.2 Evaluation of Targeted Advertising Opt-outs

Existing opt-out tools for targeted advertising include third-party cookie blockers built into web browsers, browser extensions, and opt-out tools provided by industry self-regulatory groups. The effectiveness of these tools varies. Many opt-out options, for example, prevent tailored ads from being displayed but do not opt users out of web tracking [8]. A 2012 study found certain browser extensions and cookie-based tools to be helpful in limiting targeted text-based ads, but the "Do Not Track" option in browsers was largely ineffective [6, 31].

Prior evaluations of targeted advertising opt-out tools have revealed numerous usability issues that can impose a heavy burden on users. For instance, using opt-out cookies is cumbersome, as these cookies need to be manually installed and updated, and may be inadvertently deleted [46]. Browser extensions partially mitigate these issues but introduce other problems. Leon et al. found in 2012 that descriptions of browser extensions were filled with jargon, and participants were not effectively prompted to change their settings when the tool interfered with websites [42]. Some of these tools have since been updated to address usability concerns. Opt-out tools offered by industry self-regulatory groups also exhibit low comprehension, as studies have found that the NAI's description of opt-out cookies led to the misinterpretation that the opt-out would stop all data collection by online advertisers, and DAA's AdChoices icon failed to communicate to web users that a displayed ad is targeted [48, 60]. Moreover, when the AdChoices icon is presented on a mobile device, it tends to be difficult for people to see [33].

Furthermore, studies have identified issues related to noncompliance with self-regulatory guidelines for targeted advertising. Hernandez et al. found in 2011 that among Alexa's US top 500 websites only about 10% of third-party ads used

the AdChoices icon, and even fewer used the related text [35]. Similar noncompliance issues with the enhanced notice requirement were found by Komanduri et al. in a large-scale examination of DAA and NAI members [40]. In 2015, Cranor et al. reported that privacy policies of companies who use targeted advertising did not meet self-regulatory guidelines related to transparency and linking to personally identifiable information [16]. Our analysis complements this prior work by further highlighting practices used by websites that could make advertising opt-outs difficult to use or comprehend.

#### 3.1.3 Evaluation of Data Deletion Choices

Comparatively, there have been fewer evaluations of data deletion mechanisms, likely due to the recency of corresponding legal requirements. The Global Privacy Enforcement Network (GPEN) reported that only half of the websites and mobile apps they evaluated provided instructions for removing personal data from the company's database in the privacy policy, and only 22% specified the retention time of inactive accounts [34]. An encouraging effort is the JustDelete.me database,[1] which rated the account deletion process of 511 web services. More than half of the websites analyzed (54%) were rated as having an "easy" process for deleting an account from the website. Yet, these ratings only apply to the specific action required to use deletion mechanisms and do not systematically analyze the full end-to-end interaction, which also includes finding and learning available mechanisms and assessing the result of the action, as we do in our study.

### 3.2 Programmatic Privacy Choice Extraction

Recent efforts in analyzing opt-out mechanisms have utilized automated extraction tools and machine learning. Such tools have been used to evaluate the privacy policies of US financial institutions [17] and descriptions of third-party data collection in website privacy policies [43]. Machine learning classifiers developed by Liu et al. have successfully been used to annotate privacy policy text for certain practices [45]. More directly related to privacy choice mechanisms, Sathyendra et al. and Wilson et al. developed classifiers to identify opt-out choices and deletion options in the privacy policies of websites and mobile apps [58, 62]. Ultimately, these techniques demonstrate the prospect of building tools to extract privacy choices buried in the long text of privacy policies to present them in a more user-friendly manner. However, our manual in-depth analysis of how these choices are presented by websites can identify issues and inform the design of consent mechanisms that better meet users' needs.

[1] https://backgroundchecks.org/justdeleteme/

### 3.3 Consumer Attitudes and Behavior

Prior studies have shown that consumers are uncomfortable with certain data handling practices commonly used by websites. For example, in a survey conducted by Business Week and Harris Poll in 2000, 78% of respondents were concerned that companies would use their information to send junk emails [9]. Similarly, in another 1999 survey, 70% of respondents wanted to have the choice to be removed from a website's mailing list [18]. More recently, Murillo et al. examined users' expectations of online data deletion mechanisms and found that users' reasons for deleting data were varied and largely depended on the type of service, posing difficulties for a uniform deletion interface adaptable for all services [51].

Most prior work on consumer attitudes and behavior in this area has focused on targeted advertising practices. Internet users consider targeted advertising a double-edged sword: targeted advertising stimulates purchases and is favored by consumers when it is perceived to be personally relevant; yet, it also raises significant privacy concerns due to the large amount of personal data being collected, shared, and used in a nontransparent way [7, 39]. Prior research has shown rich evidence of consumers' objection to data collection for targeted advertising purposes. In Turow et al.'s 2009 national survey, over 70% of respondents reported that they did not want marketers to collect their data and deliver ads, discounts, or news based on their interests [59]. Similarly, in McDonald and Cranor's 2010 survey, 55% of respondents preferred not to see interest-based ads, and many were unaware that opt-out mechanisms existed [48]. These findings are supported by qualitative work, such as Ur et al.'s 2012 interview study in which participants generally objected to being tracked [60].

Despite significant privacy concerns, consumers struggle to protect their online privacy against targeted advertising for multiple reasons [14, 42]. Two aspects that limit users' capabilities in dealing with targeted advertising include the asymmetric power held by entities in the targeted advertising ecosystem, and consumers' bounded rationality and limited technical knowledge to fully understand and utilize privacy-enhancing technologies [1, 3, 24]. For example, many consumers may not know that ads they see may be based on their email content [48]. Yao et al. showed that mental models about targeted advertising practices contain misconceptions, including conceptualizing trackers as viruses and speculating that trackers access local files and reside locally on one's computer [63]. These findings highlight the importance of improving the usability of opt-out tools and disclosures of data handling practices, as well as enhancing consumer education.

## 4 Methodology

We developed an analysis template for the systematic analysis of data deletion, email, and targeted advertising choices offered by websites along multiple metrics. Our analysis in-

cluded websites sampled across different ranges of web traffic that were registered primarily in the United States.

## 4.1 Template for Analysis

We implemented a comprehensive template in Qualtrics to facilitate standardized recording of data for researchers' manual content analysis of websites. For the purpose of our analysis, we defined opt-outs for email communications as mechanisms that allow users to request that a website stop sending them any type of email message (e.g., marketing, surveys, newsletters). Any mention of an advertising industry website or opt-out tool, as well as descriptions of advertising-related settings implemented by the website, browser, or operating system (e.g., "Limit Ad Tracking" in iOS) was considered as an opt-out for targeted advertising. We identified data deletion mechanisms as a means through which users can delete their account or information related to their account, including via an email to the company.

In completing the template, a member of the research team visited the home page, privacy policy, and account settings of each website examined, and answered the relevant template questions according to the privacy choices available. For each choice identified, we recorded where the privacy choice is located on the website, the user actions required in the shortest path to exercise the choice, and other information about the choice provided by the website. To complete the template, researchers were asked to:

1. Visit the homepage of the website.
2. Note if there was a notice to consumers regarding the use of cookies on the website.
3. Create a user account for the website using an alias and email address provisioned for this analysis.
4. Review any targeted advertising opt-outs on a page linked from the homepage that describes advertising practices (i.e., an "AdChoices" page).
5. Visit the website's privacy policy.
6. Review any email communications in the privacy policy.
7. Review any targeted advertising opt-outs in the policy.
8. Review any data deletion mechanisms in the policy.
9. Note whether the privacy policy mentions Do Not Track.
10. Note any other privacy choices in the privacy policy and linked pages providing privacy information.
11. Review any email communications opt-outs in the user account settings.
12. Review any targeted advertising opt-outs in the user account settings.
13. Review any data deletion mechanisms in the user account settings.
14. Note any other privacy choices in the account settings.

At every stage, researchers also made note of practices for offering privacy controls that seemed particularly detrimental or beneficial to usability throughout the Interaction Cycle, a framework for describing the end-to-end interaction between a human and a system [5].

To refine the template, our research team conducted six rounds of pilot testing with 25 unique websites from Amazon Alexa's[2] ranking of top 50 US websites. For every round of piloting, two researchers independently analyzed a small set of websites. We then reconciled disagreements in our analysis, and collaboratively revised the questions in the template to ensure that there was a mutual understanding of the metrics being collected.

## 4.2 Website Sample

We examined 150 websites sampled from Alexa's ranking of global top 10,000 websites (as of March 22, 2018). To understand how privacy choices vary across a broad range of websites, we categorized these websites based on their reach (per million users), an indicator of how popular a website is, provided by the Alexa API. We selected two thresholds to divide websites and categorized them as: *top websites* (ranks 1 - 200), *middle websites* (ranks 201 - 5,000), and *bottom websites* (ranks > 5,000). These thresholds were identified by plotting websites' reach against their rank, and observing the first two ranks at which reach leveled off. Our analysis included 50 *top*, 50 *middle*, and 50 *bottom* websites randomly selected from each range. We stratified our sample as such, since consumers may spend significant time on websites in the long tail of popularity. The stratified sample enables us to understand the privacy choices provided on low-traffic websites, and how they differ from choices on popular websites.

The ICANN "WHOIS" record of 93 websites in our sample indicated registration in the United States, while other websites were registered in Europe (26), Asia (11), Africa (4), Central America/the Caribbean (2), or contained no country related information (14). In constructing our sample, we excluded porn websites to prevent researchers' exposure to adult content. To simplify our data collection, we also excluded a handful of websites drawn during our sampling that required a non-email based verification step, or sensitive information like a social security number (SSN) or credit card, to create a user account. Due to the language competencies of the research team, we only included websites written in English, or those with English versions available. All websites included in our study were analyzed between April and October 2018. Data collected from our pilot rounds are not included in our analysis. The types of websites included in our sample ranged from popular news and e-commerce websites to university and gaming websites.

Due to the GDPR, many websites were releasing new versions of their privacy policies during the period of our data analysis. In October 2018 we reviewed all websites in our dataset that had been analyzed prior to May 25, 2018, the GDPR effective date, and conducted our analysis again on

[2] Amazon Alexa Top Sites: https://www.alexa.com/topsites

the 37 websites that had updated their privacy policy. Our reported findings are primarily based on the later versions of these policies, but we also compared the pre- and post-GDPR versions for these websites, and highlight differences.

### 4.3 Data Collection

The researchers involved in data collection went through a training process during which they completed the template for several websites prior to contributing to the actual dataset. To ensure thorough and consistent analysis, two researchers independently analyzed the same 75 (50%) websites sampled evenly across categories. Cohen's Kappa ($\kappa = 0.82$) was averaged over the questions in which researchers indicated whether or not privacy choice mechanisms were present on the page being analyzed. All disagreements in the analysis were reviewed and reconciled, and the remaining 75 websites were coded by only one researcher. Analyzing one website took 5 to 58 minutes, with an average of 21 minutes spent per website. This variance in analysis time was related to websites' practices. For example, websites that did not use email marketing or targeted advertising could be reviewed more quickly. To prevent browser cookies, cookie settings, or browser extensions from affecting website content, researchers collected data in Google Chrome's private browsing mode, opening a new browser window for each website.

### 4.4 Limitations

The privacy choices we reviewed may not be representative of all websites. Our sample only included English-language websites, which may not be reflective of websites in other languages. We also only included websites from Alexa's top 10,000 list. Websites with lower rankings may exhibit a different distribution of choices than that observed in our sample. Moreover, in the process of random sampling, we excluded a small number of websites, primarily for financial institutions, that required sensitive personal information (e.g., SSN or credit card) for account registration. Considering the sensitive nature of this type of personal information, these websites may offer privacy choices through different means or offer other choices. However, our sample still includes many websites that collect credit card information and other sensitive personal information, but do not require it for account creation. Despite these exclusions, we are confident the websites we analyzed provide broad coverage of websites' most prominent practices for offering opt-outs and deletion mechanisms.

Additionally, since our analysis was conducted using US IP addresses, we may not have observed privacy choices available to residents of other jurisdictions (such as the EU) with other legal privacy requirements. Our analysis thus only reflects privacy choices available to US-based consumers.

Lastly, our study cannot provide definite conclusions about how consumers will comprehend and utilize the privacy choices we analyzed. We chose a content analysis approach in order to be able to gain a systematic overview of current practices in provisioning opt-out choices, which was not provided by prior work at this scale. Nonetheless, based on prior opt-out evaluations and design best practices, we hypothesize that certain design choices (e.g., multiple steps to an opt-out choice) will appear difficult or confusing to users. Our findings also surface many other issues that pose challenges to consistent privacy choice design. The effects of these issues on consumers could be studied in future work.

## 5 Results

Our manual content analysis of 150 websites revealed that privacy choices are commonly available, but might be difficult to find and to comprehend. We identified several factors that likely negatively impact the usability of privacy choices, such as inconsistent placement, vague descriptions in privacy policies, and technical errors.

### 5.1 Overview of Privacy Policies

Nearly all of the websites in our sample included a link to a privacy policy from the home page. The only websites that did not include a privacy policy were three bottom websites. Of the 147 policies analyzed, 15% (22) were a corporate policy from a parent company. In line with prior findings, comprehension of the text that describes privacy choices requires advanced reading skills [26]. However, about a third of policies in our analysis adopted tables of contents to present the information in a structured way, or linked to separate pages to highlight particular sections of the policy.

**Privacy choices text has poor readability.** For websites in our sample that had a privacy policy, we recorded the policy text and marked out the portions that described privacy choices. We then conducted a readability analysis using the text analysis service readable.io.

As reported in Table 1, the Flesch Reading Ease Scores (FRES) for text related to email opt-outs, targeted advertising opt-outs, and data deletion choices received means and medians of about 40 on a 0 to 100 point scale (with higher scores indicating easier-to-read text) [32]. The analyzed text for all three types of privacy choices on the Flesch-Kincaid Grade Level (FGL), a grade-based metric, had means and medians around 13, which implies the text requires the audience to have university-level reading abilities. On Flesch's 7-level ranking system, over 90% of the analyzed privacy choices were described in text that was "very difficult," "difficult," or "fairly difficult" to read.

Privacy policies as a whole had better, but not ideal, readability, compared to privacy choice text: our analyzed privacy

| | Flesch Reading Ease | | Flesch-Kincaid | |
|---|---|---|---|---|
| | Mean | SD | Mean | SD |
| Email Comm. | 39.54 | 13.55 | 13.89 | 3.40 |
| Targeted Adv. | 39.38 | 15.41 | 13.72 | 4.48 |
| Data Deletion | 38.98 | 17.89 | 14.28 | 5.40 |
| Privacy Policies | 45.80 | 10.72 | 10.20 | 2.44 |

Table 1: Readability scores for privacy policy text describing email opt-outs, advertising opt-outs, and deletion choices.

policies had a mean FRES of 45.80 and a mean FGL of 10.20, which align with prior readability evaluations of privacy policies, both across domains [26] and for particular categories (e.g., social networking, e-commerce, and healthcare websites [23, 49]). Nevertheless, literacy research suggests materials approachable by the general public should aim for a junior high reading level (i.e., 7 to 9) [36]. These statistics of our analyzed privacy policies and text related to privacy choices, which were all post-GDPR versions, suggest that most of them still fail to comply with the GDPR's "clear and plain language" requirement, a key principle of transparency.

**Some websites use table of contents and support pages.** We also observed that a significant portion of the policies in our sample were organized using a table of contents. Of the 147 privacy policies, 48 (33%) included a table of contents, which provides a road map for users to navigate a policy's sections. Additionally, 53 (36%) policies linked to secondary pages related to the company's privacy practices. For example, Amazon and Dropbox have individual pages to explain how targeted advertising works and how to opt-out.

## 5.2 Presence of Privacy Choices

In this section, we first focus on whether and where choices were present on the websites analyzed. More details about how these choices are described in policies are presented in Section 5.3. We found that privacy choices are commonly offered across all three website tiers. Beyond privacy policies, websites often provide opt-outs and data deletion choices through other mechanisms, such as account settings or email.

**Privacy choices are prevalent.** All three types of privacy choices were prevalent in our sample. As seen in Table 2, 89% of websites with email marketing or targeted advertising offered opt-outs for those practices, and 74% of all websites had at least one data deletion mechanism. The location of privacy choices across top, middle, and bottom websites is displayed in Figure 1. Top websites were found to provide more privacy choices than middle and bottom websites.

| | Email Comm. | Targeted Adv. | Data Deletion |
|---|---|---|---|
| # of sites applicable | 112 | 95 | 150 |
| # of sites choice present | 100 | 85 | 111 |
| % of applicable sites | 89% | 89% | 74% |

Table 2: Summary of the availability of each type of privacy choice and websites on which they are applicable.

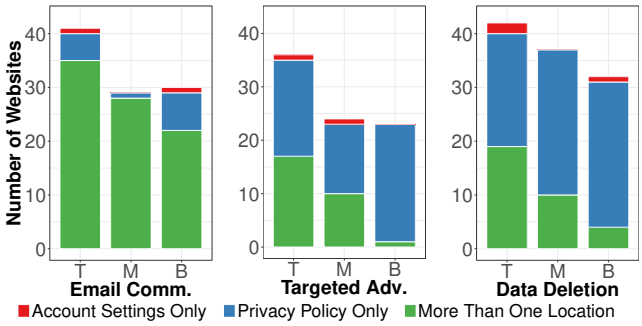



Figure 1: Location of privacy choices for top, middle, and bottom websites. Top websites offered the most privacy choices.

**Email opt-outs were links in policies and emails.** Most often, opt-outs for email communications were offered in multiple ways. Nearly all (98 of 100) websites offering email communication opt-outs presented the opt-out for emails in the privacy policy; however, only 31 policies included a direct link to the opt-out page, while 70 stated that users could unsubscribe within emails. Additionally, 51 websites had an opt-out in the account settings, the majority of which (33) lead to the same opt-out described in the privacy policy, and 15 websites provided a choice for email communication during account creation.

**Advertising opt-outs were links in privacy policies.** Websites primarily used their privacy policy to provide opt-outs for targeted advertising. Of 85 websites that offer at least one targeted advertising opt-out, 80 provided them in the privacy policy. Among them, 74 also provided at least one link, while the remaining just described an opt-out mechanism with text, such as "...you can opt out by visiting the Network Advertising initiative opt out page." However, 58 websites had multiple links leading to different opt-out tools, which may cause confusion about which tool visitors should prioritize and what the differences are.

On 26 websites, an "AdChoices" page linked from the homepage described the website's advertising practices and presented opt-out choices. Among them, 15 used text containing the words "ad choices" to refer to the page; others labeled the page as "interest-based ads," "cookie information" or "cookie policy." Additionally, 12 websites included opt-

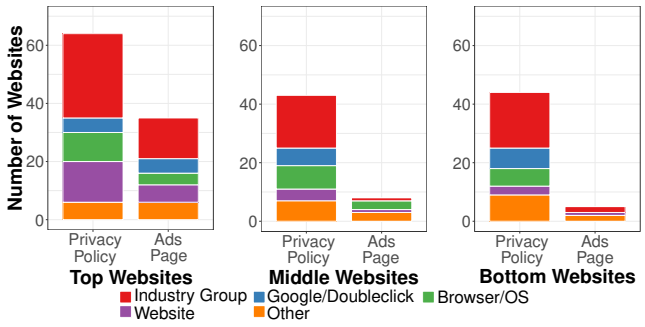



Figure 2: Distribution of different types of targeted advertising opt-outs in privacy policies and "About Ads" pages across top, middle, and bottom websites.

outs in the user account settings, 11 of which led to the same opt-out page presented in the policy.

As seen in Figure 2, many websites referred to opt-out tools provided by advertising industry associations. However, 27% of opt-out links pointing to the DAA or NAI directed visitors to their homepages, instead of their opt-out tools. This creates a substantial barrier for people to opt-out because visitors still need to find the appropriate opt-out tool on the DAA and NAI websites. Conversely, 21 of 22 links to the European Interactive Digital Advertising Alliance (EDAA) in the website policies led directly to the EDAA's opt-out tool. Less common, some websites provided advertising opt-outs implemented by Google or the website itself. Others provided instructions for adjusting cookie or ad related settings in the browser or operating system, such as the "Limit Ad Tracking" setting in iOS. The use of other services like TrustArc (formerly TRUSTe) or Evidon was also relatively rare.

**Data deletion controls were provided in privacy policies and account settings.** We observed that 111 websites in our sample (74%) provided data deletion mechanisms to their users, which is higher than the 51% in the sample analyzed by GPEN in 2017 [34]. Among websites offering deletion mechanisms, 75 only provided the choices through the privacy policy, three only displayed them in the user account settings, and 33 provided them through multiple locations. However, even when data deletion choices are described in the privacy policy, only 27 policies included a direct link to a data deletion tool or request form. The more common practice was to offer instructions about how to email a data deletion request, as was done in 81 policies.

**The GDPR contributed to more deletion controls.** In our sample, 37 websites updated their privacy policy around the GDPR effective date. Four websites added their privacy policies post-GDPR. Most of the 37 websites had already included descriptions of privacy choices before the GDPR effective date, especially for marketing opt-outs (29 out of 37). In our sample, the GDPR had the greatest impact on data deletion controls, with 13 websites adding instructions for deleting account data to their post-GDPR privacy policy. However, such dramatic change was not observed for marketing and targeted advertising opt-outs.

**Websites include other data collection controls.** Though less common, some websites described additional privacy-related opt-outs in their privacy policy and account settings. Opt-outs for web analytic services (e.g., Google Analytics) were offered by 21% (31) of websites. Interestingly, 17 websites offered opt-outs for the sharing of personal information with third parties. For example. CNN's privacy policy[3] stated that "We may share the Information with unaffiliated Partners and third parties..." and provided a link to an opt-out from such sharing. Additionally, nine websites described controls offered by the website, browser, or operating system related to the use of location history or location data.

Only 28 of the 150 websites analyzed (19%) displayed a cookie consent notice on their home page, alerting users that cookies are being used on the website and getting consent to place cookies in the user's browser. Among them, only five offered a means to opt out or change cookie related settings. However, as these websites were accessed from US IP addresses, we may have observed different practices than those offered to EU-based visitors. Prior work has found a substantial increase in cookie consent notices on European websites post-GDPR [20].

**Do Not Track has low adoption.** Of the 150 websites analyzed, only eight (5%) specified that they would honor Do Not Track (DNT), a mechanism that allows users to express that they wish not to be tracked by websites, while 48 (32%) explicitly stated that the website will not honor it [31]. Another 91 (61%) did not specify whether or not they would respect the DNT header, which is in violation of the California Online Privacy Protection Act (CalOPPA) [10].

## 5.3 Descriptions of Choices in Privacy Policies

In addition to analyzing whether privacy choices are present in privacy policies, we analyzed *how* those choices are presented or described. We found a lack of consensus in the wordings used to present privacy choices. Additionally, many websites provided little information regarding what actually happened when a targeted advertising opt-out or data deletion choice was exercised, thus potentially confusing or misleading users.

**There is no dominant wording for section headings.** Table 3 summarizes common bigrams and trigrams in policy section headings related to privacy choices. Across policies,

[3] https://www.cnn.com/privacy

| N-Gram | Email Comm. | Targeted Adv. | Data Deletion |
|---|---|---|---|
| how we use | 9 | 5 | 2 |
| opt out | 13 | 7 | 2 |
| person* data | 8 | 1 | 10 |
| person* inform* | 7 | 2 | 13 |
| third part* | 0 | 14 | 2 |
| we collect | 15 | 7 | 5 |
| we use | 11 | 5 | 2 |
| your choic* | 11 | 9 | 10 |
| your inform* | 7 | 3 | 10 |
| your right* | 9 | 2 | 20 |

Table 3: Bigrams and trigrams occurring in at least 5% of privacy policy section headings. Counts are the number of policies (out of 147) in which a n-gram occurred in the headings of sections containing a privacy choice. Some policies described the same privacy choice under multiple headings, or used multiple n-grams in a heading.

similar headings were used to present all three types of privacy choices, e.g., referring to collection and use of personal data or information, or describing a visitor's rights or choices. In contrast, the bigram "opt out" more commonly referred to choices related to email communications or targeted advertising. Similarly, advertising opt-outs were sometimes presented under sections describing third parties, which is not as applicable to the other two types of privacy choices. However, no single n-gram occurred in more than 20 of the policies we analyzed. This lack of consistency across websites could make locating privacy choices across websites difficult for visitors. Furthermore, some policies included multiple headings related to privacy choices, which could also potentially add significant burden to visitors.

**Most marketing opt-outs are first-party.** Among the 98 websites that provided at least one marketing communication opt-out in their privacy policy, 80 websites offered opt-outs from the website's own marketing or promotions. Additionally, 20 policies stated it is possible to opt out of marketing or promotions from third-party companies, and 19 policies specified that visitors could opt out of receiving website announcements and updates. Other less common forms of emails sent by websites that could be opted out from included newsletters, notifications about user activity, and surveys. Some websites offered opt-outs for different types of communications, such as SMS communications (10) and phone calls (8).

**Targeted advertising opt-outs are ambiguous.** We observed that privacy policies typically did not describe whether visitors were opting out of tracking entirely or just the display of targeted ads. Only 39 of the 80 websites that offered opt-outs for targeted advertising within their privacy policy made this distinction within the policy text. Among them, 32 websites explicitly stated that the opt-out only applied to the *display* of targeted ads. This lack of distinction could be confusing to visitors who desire to opt-out of *tracking* on the websites for targeted advertising purposes.

The same ambiguity exists with respect to whether an opt-out applies across multiple browsers and devices. Seventy-three websites' policies did not specify whether the opt-out would be effective across different devices, and 72 did not clarify whether the opt-out applied across all the browsers a visitor uses.

**Data deletion mechanisms vary by website.** The data deletion mechanisms presented in the privacy policies of 108 websites varied. Visitors had the option to select certain types of information to be removed from their account on 80 websites. Furthermore, 41 websites offered the option to have the account permanently deleted, and 13 allowed visitors to temporarily suspend or deactivate their account.

How soon the data would actually be deleted was often ambiguous. Ninety of 108 websites offering deletion did not describe a time frame in which a user's account would be permanently deleted and only four policies stated that information related to the account would be deleted "immediately." Another three claimed the time frame to be 30 days, and two websites said the deletion process could take up to one year.

## 5.4 Usability of Privacy Choices

Our analysis included how many steps visitors had to take to exercise a privacy choice. We found that email communications opt-outs, on average, required the most effort. We also recorded specific usability issues on 71 websites (30 top, 23 middle, and 18 bottom) that could make privacy choices difficult or impossible to use, such as missing information and broken links.

**Privacy choices require several user actions.** We counted user actions as the number of clicks, hovers, form fields, radio buttons, or check boxes encountered from a website's home page up until the point of applying the privacy choice. Table 4 displays summary statistics related to the shortest path available to exercise choices of each type. Opt-outs for email communications and data deletion choices, on average, contained more user actions, particularly check boxes and form elements, compared to opt-outs for targeted advertising. This is likely due to the reliance on the DAA and NAI opt-out tools, which typically required two or three clicks to launch the tool. Data deletion and email communications choices, on the other hand, often required form fields or additional confirmations. At the extreme end, 38 user actions were required to complete the New York Times' data deletion request form, which included navigating to the privacy policy, following the link to the request form, selecting a request type, selecting up

| | Clicks | Boxes | Hovers | Form | Other | Total |
|---|---|---|---|---|---|---|
| Email Comm. | 2.90 | 1.68 | 0.38 | 0.33 | 0.17 | **5.32** |
| Targeted Adv. | 2.80 | 0.10 | 0.25 | 0.00 | 0.01 | **3.16** |
| Data Deletion | 2.93 | 1.05 | 0.23 | 1.07 | 0.05 | **5.32** |

Table 4: Average number of actions required in the shortest path to exercise privacy choices, counted from the home page up until, but not including, the action recording the choice (i.e., "save/apply" button).

to 22 check boxes corresponding to different New York Times services, filling in eight form fields, selecting four additional confirmation boxes, and completing a reCAPTCHA.[4]

**Policies contain missing, misleading, or unhelpful information.** Many choice mechanisms were confusing or impossible to use because of statements in the website's privacy policy. In six instances, text in the policy referred to an opt-out, but that opt-out did not exist or the website did not provide vital information, such as an email address to which visitors can send privacy requests. Six websites included misleading information in the policy text, such as presenting the Google Analytics opt-out browser extension as an opt-out for targeted advertising,[5] and omitting mentions of targeted advertising in the privacy policy while providing opt-outs elsewhere on the website. Additionally, seven websites mentioned user accounts in the privacy policy but no mechanisms to create a user account were observed on the website. Two of these cases were TrustedReviews and Space.com, whose policies covered multiple domains, including some with user accounts. These issues appeared in fairly equal frequency across top, middle, and bottom websites.

**Some websites had broken choice mechanisms and links.** We also recorded 15 instances in which provided links to relevant privacy choice information or mechanisms were broken or directed to an inappropriate location, such as the website's homepage, or the account settings for a parent website. We further observed that four websites offered choice mechanisms that did not appear to properly function. For example, on Rolling Stone's email preferences page, selections made by visitors seemed to be cleared on every visit. GamePress's data deletion request form was implemented by Termly and did not seem to refer to GamePress, making it unclear where and how the form would be processed.

**Some websites made poor design choices.** We noted several website design choices that may impact the usability of privacy choices. On ten websites, we observed a privacy policy displayed in an unconventional format, such as in a PDF or in a modal pop-up dialogue, instead of a normal HTML page. This may impact how well visitors can search for privacy choices in a policy. Another design choice that impacted searchability was collapsing the policy text under section headings; keyword search is not effective unless all sections are opened. Five policies also had stylistic issues with their policies, such as including opt-out links that were not clickable or advertisements in the middle of the policy. Some websites offered burdensome pages for managing email communication settings, requiring visitors to individually deselect each type of communication sent by the website. Others placed the option for opting out of all communications *after* a long list of different types of content, rather than before it, making it less visible. For example, Amazon offered this option after listing 79 different communications, which rendered it invisible until scrolling much further down the page.

### 5.4.1 Aids for privacy choice expression

Conversely, a few websites made additional efforts to make their privacy choices more accessible to visitors. Many opt-outs (such as the Google Ad Settings page) went into effect once a visitor expressed a privacy choice, and did not require the additional step of pressing a confirmation (i.e., "save/apply"). Some, like Metacrawler, centralized the privacy choices related to email communications, targeted advertising, and data deletion into a single section of the policy. Others, including Fronter, were diligent about providing links to related privacy information, such as regulation or the privacy policies of third parties used by the website. To further aid visitors, three websites (BBC, Garena, and LDOCE Online) presented important privacy information in a "Frequently Asked Questions" format. Moreover, Google and Booking.com, provided users with a short video introducing their privacy practices.

## 6 Improving Privacy Choices

Our findings indicate that certain design decisions may make exercising privacy choices difficult or confusing, and potentially render these choices ineffective. We provide several *design* and *policy* recommendations for improving the usability of web privacy choices. Our recommendations not only serve as concrete guidelines for website designers and engineers, but also have the potential to help policy makers understand current opt-out practices, their deficiencies, and areas for improvement. These suggestions could then be integrated into future guidelines, laws, and regulations.

Our discussion is based on the Interaction Cycle, which divides human interaction with systems into four discrete stages [5]. It serves as a framework to highlight the cognitive and physical processes required to use choice mechanisms,

[4] reCAPTCHA: https://www.google.com/recaptcha/intro/v3.html

[5] Google merged its advertising and analytics platforms in July 2018, but the Google Analytics opt-out extension only pertains to analytics tracking.

and in turn synthesizes our findings to address specific usability barriers. We mapped the expression of online privacy choices to the Interaction Cycle as: 1) finding, 2) learning, 3) using, and 4) understanding a privacy choice mechanism.

## 6.1 Finding Privacy Choices

**Use standardized terminology in privacy policies.** As noted in Section 5.3, no single n-gram was present in an overwhelming majority of privacy policy section headings in which choices were described, and there was much variation in how websites offered privacy choices. For example, data deletion mechanisms were placed under headings like "What do you do if you want to correct or delete your personal information?" in some policies, but under more general headings like "Your Rights" in others. Even more confusing, some policies contained multiple titles similar to both of these.

Inconsistencies across different privacy policies may make finding specific privacy choices difficult. We recommend that future privacy regulations include requirements for standardized privacy policy section headings. Such guidance exists for privacy notices of financial institutions in the United States, as well as data breach notifications to California residents [11, 61]. Our results highlight the most common terms that websites already use in providing privacy choices, which could serve as a foundation for formulating such guidance.

**Unify choices in a centralized location.** Websites sometimes offer different opt-out choices on different pages of the website for the same opt-out type. This problem is most salient for targeted advertising opt-outs, which could appear either in privacy policies, account settings, or an individual "AdChoices" page linked to from the home page. Furthermore, some privacy policies did not link to the "AdChoices" page or the account settings where the advertising opt-outs were located. Therefore, by looking at just the privacy policy, which may be where many users would expect to find privacy choices, visitors would miss these opt-outs available to them.

One potential solution is having all types of privacy choices in a centralized location. This can be achieved as a dedicated section in the privacy policy, or even as an individual page with a conspicuous link provided on the home page. However, it will likely require regulatory action for many companies to prioritize reorganizing their current opt-outs in this way.

## 6.2 Learning How To Use Privacy Choices

**Simplify or remove decisions from the process.** Another practice that adds to the complexity of exercising opt-outs is the presence of links to multiple tools. For instance, more than one third (58) of our analyzed websites provided links to multiple advertising opt-outs. To simplify the privacy choice process, websites should unify multiple choice mechanisms into a single interface, or provide one single mechanism for a particular type of privacy choice. If not technically feasible, websites should help visitors distinguish the choices offered by each mechanism.

**Ensure all choices in the policy are relevant.** The use of one policy for a family of websites might be the reason for some of the points of confusion highlighted in Section 5.4. These corporate "umbrella policies" might explain cases where we observed links from the privacy policy directing to unrelated pages on a parent company's website, or references to account settings even when the website does not offer mechanisms to create user accounts. While maintaining one policy may be easier for parent companies, this places a substantial burden on visitors to identify the practices that apply to a particular website.

To mitigate such issues, companies should carefully check if the information provided in the privacy policy matches the websites' actual practices. If an umbrella policy is used across multiple websites, practices should be clearly labelled with the websites to which they are applicable. Regulatory authorities should further exert pressure by emphasizing the necessity of having accurate privacy policies and conducting investigations into compliance.

## 6.3 Using Privacy Choices

**Simplify multi-step processes.** We noted that privacy choices typically require multiple steps, which may frustrate and confuse users. As described in Section 5.4, our analyzed privacy choices required an average of three to five user actions prior to pressing a button to apply the choice, assuming the visitor knew which pages to navigate to in advance. On the extreme end, completing one deletion request form required 38 user actions, as the interface included several boxes related to different services offered by the website. Though this type of interface allows users to have greater control, websites should also have a prominent "one-click" opt-out box available to visitors.

It is also conceivable that many companies may deliberately make using privacy choices difficult for their visitors. In this case, it is up to regulators to combat such "dark patterns." [2, 54] Though it may be unrealistic to set a threshold for the maximum number of user actions required to exercise a privacy choice, regulators should identify websites where these processes are clearly purposefully burdensome and take action against these companies. This would both serve as a deterrent to other companies and provide negative examples. Precedents of such regulatory action have emerged, such as a ruling by the French Data Protection Authority (the "CNIL") which found that Google fails to comply with the GDPR's transparency requirement as its mobile phone users need "up to five or six actions to obtain the relevant information about the data processing" when creating a Google account [37].

Some of our analyzed websites have already provided exemplary practices to simplify privacy choices, e.g., automatically applying privacy choices once the user selects or deselects an option, rather than requiring the user to click an additional "save" or "apply" button. Clicking an additional button may not be intuitive to users, especially if it is not visible without scrolling down the page. Removing this extra step would avoid post-completion errors, in which a user thinks they have completed privacy choice, but their choice is not registered by the website. A requirement that all changes in privacy settings must be automatically saved could be integrated into regulations and related guidelines. However, any changes should be made clear to the user to avoid accidental changes.

**Provide actionable links.** Our findings show that the use of links pointing to privacy choices was not ubiquitous, and varied substantially across different types of privacy choices; 93% of websites that offered the choice to opt out of targeted advertising provided at least one link, whereas the percentage for email communication opt-out and data deletion choice was 32% and 24% respectively. Websites that do not provide links usually provide text explanations for the opt-out mechanisms instead. However, visitors may not follow the text instructions if significant effort is required, such as checking promotional emails in their personal inbox for the "unsubscribe" link, or sending an email to request their account to be deleted. We also found that some websites may not provide sufficient guidance to support exercising a privacy choice.

Our findings point to the necessity to enhance the actionability of privacy choices by providing links. However, there should be a careful decision about how many links to include and where to place them. Ideally, only one link for one particular type of opt-out should be provided. When multiple links are presented on the same page, there needs to be sufficient contextual information to help users distinguish these links. Of equal importance is the functionality of provided links. In our analysis, we observed a few instances in which the provided links were broken, directed to an inappropriate location, or had styling that easily blended in with text. These practices reduce the actionability of the corresponding privacy choice and negatively impact the user experience.

## 6.4 Understanding Privacy Choices

**Describe what choices do.** We found that privacy policies did not provide many details that informed visitors about what a privacy choice did, particularly in the cases of targeted advertising opt-outs and data deletion choices. Among all websites that provided targeted advertising opt-outs, fewer than 15% distinguished opting out of tracking from opting out of the display of targeted ads, or indicated whether the opt-out was effective on just that device or browser or across all their devices and browsers. Similarly, among all websites that provided data deletion choices, only 19% stated a time frame for when the account would be permanently deleted.

Future regulations could stipulate aspects that must be specified when certain opt-outs are provided (e.g., the device that the opt-out applies to). This may reduce instances where visitors form expectations that are misaligned with a companies' actual practices.

# 7 Conclusion

We conducted an in-depth empirical analysis of data deletion mechanisms and opt-outs for email communications and targeted advertising available to US consumers on 150 websites sampled across three ranges of web traffic. It is encouraging that opt-outs for email communications and targeted advertising were present on the majority of websites that used these practices, and that almost three-quarters of websites offered data deletion mechanisms. However, our analysis revealed that presence of choices is not the same as enabling visitors to execute the choice. Through our holistic content analysis, we identified several issues that may make it difficult for visitors to find or exercise their choices, including broken links and inconsistent placement of choices within policies. Moreover, some policy text describing choices is potentially misleading or likely does not provide visitors with enough information to act. Design decisions may also impact the ability of visitors to find and exercise available opt-outs and deletion mechanisms. We offer several design and policy suggestions that could improve the ability of consumers to use consent and privacy control mechanisms.

# Acknowledgments

This project is funded by the National Science Foundation under grants CNS-1330596 and CNS-1330214. We wish to acknowledge all members of the Usable Privacy Policy Project (www.usableprivacy.org) for their contributions.

# References

[1] Alessandro Acquisti. Privacy in electronic commerce and the economics of immediate gratification. In *Proceedings of the Conference on Electronic Commerce (EC)*, pages 21–29, 2004.

[2] Alessandro Acquisti, Idris Adjerid, Rebecca Balebako, Laura Brandimarte, Lorrie Faith Cranor, Saranga Komanduri, Pedro Giovanni Leon, Norman Sadeh, Florian Schaub, Manya Sleeper, et al. Nudges for privacy and security: Understanding and assisting users' choices online. *ACM Computing Surveys (CSUR)*, 50(3):44, 2017.

[3] Alessandro Acquisti and Jens Grossklags. Privacy and rationality in individual decision making. *IEEE Security & Privacy (S&P)*, 3(1):26–33, 2005.

[4] Hazim Almuhimedi, Florian Schaub, Norman Sadeh, Idris Adjerid, Alessandro Acquisti, Joshua Gluck, Lorrie Faith Cranor, and Yuvraj Agarwal. Your location has been shared 5,398 times!: A field study on mobile app privacy nudging. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, pages 787–796, 2015.

[5] Terence S Andre, H Rex Hartson, Steven M Belz, and Faith A McCreary. The user action framework: A reliable foundation for usability engineering support tools. *International Journal of Human-Computer Studies*, 54(1):107–136, 2001.

[6] Rebecca Balebako, Pedro Leon, Richard Shay, Blase Ur, Yang Wang, and Lorrie Faith Cranor. Measuring the effectiveness of privacy tools for limiting behavioral advertising. In *Proceedings of the Web 2.0 Security and Privacy Workshop (W2SP)*, 2012.

[7] Alexander Bleier and Maik Eisenbeiss. The importance of trust for personalized online advertising. *Journal of Retailing*, 91(3):390–409, 2015.

[8] Sophie C Boerman, Sanne Kruikemeier, and Frederik J Zuiderveen Borgesius. Online behavioral advertising: A literature review and research agenda. *Journal of Advertising*, 46(3):363–376, 2017.

[9] Bloomberg Businessweek. Business Week/Harris Poll: A Growing Threat. page 96, 2000.

[10] California Legislative Information. Online privacy protection act of 2003 - California business and professions code sections 22575-22579, 2003.

[11] California State Government. California civ. code s. 1798.82(a), 2003. https://leginfo.legislature.ca.gov/faces/codes_displaySection.xhtml?lawCode=CIV&sectionNum=1798.82.

[12] California State Legislature Website. SB-1121 California consumer privacy act of 2018, 2018. https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180SB1121.

[13] Josh Constine. A flaw-by-flaw guide to Facebook's new GDPR privacy changes, May 2018. https://techcrunch.com/2018/04/17/facebook-gdpr-changes/.

[14] Lorrie Faith Cranor. Can users control online behavioral advertising effectively? *IEEE Security & Privacy (S&P)*, 10(2):93–96, 2012.

[15] Lorrie Faith Cranor. Necessary but not sufficient: Standardized mechanisms for privacy notice and choice. *Journal on Telecommunications & High Technology Law*, 10:273, 2012.

[16] Lorrie Faith Cranor, Candice Hoke, Pedro Giovanni Leon, and Alyssa Au. Are they worth reading? An in-depth analysis of online trackers' privacy policies. *A Journal of Law and Policy for the Information Society*, 11:325, 2015.

[17] Lorrie Faith Cranor, Pedro Giovanni Leon, and Blase Ur. A large-scale evaluation of U.S. financial institutions' standardized privacy notices. *Transactions on the Web*, 10(3):17, 2016.

[18] Lorrie Faith Cranor, Joseph Reagle, and Mark S Ackerman. Beyond concern: Understanding net users' attitudes about online privacy. Technical report, TR 99.4.1, AT&T Labs-Research, 1999.

[19] Paresh Dave. Websites and online advertisers test limits of European privacy law, 2018. https://www.reuters.com/article/us-europe-privacy-advertising-gdpr/websites-and-online-advertisers-test-limits-of-european-privacy-law-idUSKBN1JS0GM.

[20] Martin Degeling, Christine Utz, Christopher Lentzsch, Henry Hosseini, Florian Schaub, and Thorsten Holz. We value your privacy... now take some cookies: Measuring the GDPR's impact on web privacy. In *Proceedings of Network and Distributed System Security Symposium (NDSS '19)*, 2019.

[21] Digital Advertising Alliance. Self-regulatory principles for online behavioral advertising, July 2009. http://digitaladvertisingalliance.org/principles.

[22] Electronic Frontier Foundation. Do not track. https://www.eff.org/issues/do-not-track.

[23] Tatiana Ermakova, Benjamin Fabian, and Eleonora Babina. Readability of privacy policies of healthcare websites. In *Proceedings of Wirtschaftsinformatik*, pages 1085–1099, 2015.

[24] José Estrada-Jiménez, Javier Parra-Arnau, Ana Rodríguez-Hoyos, and Jordi Forné. Online advertising: Analysis of privacy threats and protection approaches. *Computer Communications*, 100:32–51, 2017.

[25] European Commission. 2018 reform of EU data protection rules. https://ec.europa.eu/commission/

priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en.

[26] Benjamin Fabian, Tatiana Ermakova, and Tino Lentz. Large-scale readability analysis of privacy policies. In *Proceedings of the International Conference on Web Intelligence (WI)*, pages 18–25, 2017.

[27] Federal Trade Commission. Privacy online: Fair information practices in the electronic marketplace, May 2000. https://www.ftc.gov/sites/default/files/documents/reports/privacy-online-fair-information-practices-electronic-marketplace-federal-trade-commission-report/privacy2000.pdf.

[28] Federal Trade Commission. Protecting consumer privacy in an era of rapid change: Recommendations for businesses and policymakers, March 2012. https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf.

[29] Federal Trade Commission. CAN-SPAM Act: A compliance guide for business, March 2017. https://www.ftc.gov/tips-advice/business-center/guidance/can-spam-act-compliance-guide-business.

[30] Federal Trade Commission. Children's online privacy protection rule: A six-step compliance plan for your business, June 2017. https://www.ftc.gov/tips-advice/business-center/guidance/childrens-online-privacy-protection-rule-six-step-compliance.

[31] Roy T Fielding and David Singer. Tracking preference expression (DNT). W3C candidate recommendation, 2017. https://www.w3.org/TR/tracking-dnt/.

[32] Rudolf Franz Flesch. *Art of Readable Writing*. Harper, 1949.

[33] Stacia Garlach and Daniel Suthers. 'I'm supposed to see that?'AdChoices usability in the mobile environment. In *Proceedings of the Hawaii International Conference on System Sciences (HICSS)*, 2018.

[34] Global Privacy Enforcement Network. GPEN Sweep 2017: User controls over personal information, October 2017. https://www.privacyenforcement.net/sites/default/files/2017%20GPEN%20Sweep%20-%20International%20Report.pdf.

[35] J Hernandez, A Jagadeesh, and J Mayer. Tracking the trackers: The AdChoices icon, 2011. http://cyberlaw.stanford.edu/blog/2011/08/tracking-trackers-adchoices-icon.

[36] Mark Hochhauser. Lost in the fine print: Readability of financial privacy notices, July 2001. https://www.privacyrights.org/blog/lost-fine-print-readability-financial-privacy-notices-hochhauser.

[37] Hunton Andrews Kurth LLP. CNIL fines Google €50 million for alleged GDPR violations, January 2019. https://www.huntonprivacyblog.com/2019/01/23/cnil-fines-google-e50-million-for-alleged-gdpr-violations/.

[38] IAB Europe. EU framework for online behavioural advertising, April 2011. https://www.edaa.eu/wp-content/uploads/2012/10/2013-11-11-IAB-Europe-OBA-Framework_.pdf.

[39] Hyejin Kim and Jisu Huh. Perceived relevance and privacy concern regarding online behavioral advertising (OBA) and their role in consumer responses. *Journal of Current Issues & Research in Advertising*, 38(1):92–105, 2017.

[40] Saranga Komanduri, Richard Shay, Greg Norcie, and Blase Ur. AdChoices? Compliance with online behavioral advertising notice and choice requirements. *A Journal of Law and Policy for the Information Society*, 7, 2011.

[41] Neelie Kroes. Online privacy – reinforcing trust and confidence, June 2011. http://europa.eu/rapid/press-release_SPEECH-11-461_en.htm.

[42] Pedro Leon, Blase Ur, Richard Shay, Yang Wang, Rebecca Balebako, and Lorrie Faith Cranor. Why Johnny can't opt out: A usability evaluation of tools to limit online behavioral advertising. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, 2012.

[43] Timothy Libert. An automated approach to auditing disclosure of third-party data collection in website privacy policies. In *Proceedings of the World Wide Web Conference (The Web Conference)*, pages 207–216, 2018.

[44] Thomas Linden, Hamza Harkous, and Kassem Fawaz. The privacy policy landscape after the GDPR. *arXiv preprint arXiv:1809.08396*, 2018.

[45] Frederick Liu, Shomir Wilson, Peter Story, Sebastian Zimmeck, and Norman Sadeh. Towards automatic classification of privacy policy text. Technical report, CMU-ISR-17-118R, Carnegie Mellon University, 2018.

[46] Jonathan R Mayer and John C Mitchell. Third-party web tracking: Policy and technology. In *Proceedings of the IEEE Symposium on Security and Privacy (S&P)*, 2012.

[47] Aleecia M McDonald and Lorrie Faith Cranor. The cost of reading privacy policies. *A Journal of Law and Policy for the Information Society*, 4:543, 2008.

[48] Aleecia M McDonald and Lorrie Faith Cranor. Americans' attitudes about internet behavioral advertising practices. In *Proceedings of the Workshop on Privacy in the Electronic Society (WPES)*, 2010.

[49] Gabriele Meiselwitz. Readability assessment of policies and procedures of social networking sites. In *International Conference on Online Communities and Social Computing (OCSC)*, pages 67–75. Springer, 2013.

[50] Michael Morgan, Daniel Gottlieb, Matthew Cin, Jonathan Ende, Amy Pimentel, and Li Wang. California enacts a groundbreaking new privacy law, June 2018. https://www.mwe.com/en/thought-leadership/publications/2018/06/california-enacts-groundbreaking-new-privacy-law.

[51] Ambar Murillo, Andreas Kramm, Sebastian Schnorf, and Alexander De Luca. "If I press delete, it's gone" - User understanding of online data deletion and expiration. *Proceedings of the Symposium on Usable Privacy and Security (SOUPS)*, pages 329–339, 2018.

[52] Network Advertising Initiative. NAI code of conduct, 2018. https://www.networkadvertising.org/sites/default/files/nai_code2018.pdf.

[53] Nielsen Norman Group. Top 10 design mistakes in the unsubscribe experience, April 2018. https://www.nngroup.com/articles/unsubscribe-mistakes/.

[54] Norwegian Consumer Council. Deceived by design: How tech companies use dark patterns to discourage us from exercising our rights to privacy, June 2018. https://fil.forbrukerradet.no/wp-content/uploads/2018/06/2018-06-27-deceived-by-design-final.pdf.

[55] Online Trust Alliance. Email marketing & unsubscribe audit, December 2017. https://otalliance.org/system/files/files/initiative/documents/2017emailunsubscribeaudit.pdf.

[56] Joel R Reidenberg, N Cameron Russell, Alexander J Callen, Sophia Qasir, and Thomas B Norton. Privacy harms and the effectiveness of the notice and choice framework. *I/S: A Journal of Law and Policy for the Information Society (ISJLP)*, 11:485, 2015.

[57] John A Rothchild. Against notice and choice: The manifest failure of the proceduralist paradigm to protect privacy online (or anywhere else). *Cleveland State Law Review*, 66:559, 2017.

[58] Kanthashree Mysore Sathyendra, Shomir Wilson, Florian Schaub, Sebastian Zimmeck, and Norman Sadeh. Identifying the provision of choices in privacy policy text. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2017.

[59] Joseph Turow, Jennifer King, Chris Jay Hoofnagle, Amy Bleakley, and Michael Hennessy. Americans reject tailored advertising and three activities that enable it. 2009. https://ssrn.com/abstract=1478214.143.

[60] Blase Ur, Pedro Giovanni Leon, Lorrie Faith Cranor, Richard Shay, and Yang Wang. Smart, useful, scary, creepy: Perceptions of online behavioral advertising. In *Proceedings of the Symposium on Usable Privacy and Security (SOUPS)*, 2012.

[61] U.S. Federal Register 74. Final model privacy form under the Gramm-Leach-Bliley act, 2009.

[62] Shomir Wilson, Florian Schaub, Frederick Liu, Kanthashree Mysore Sathyendra, Daniel Smullen, Sebastian Zimmeck, Rohan Ramanath, Fei Liu, Norman Sadeh, and Noah A Smith. Analyzing privacy policies at scale: From crowdsourcing to automated annotations. *Transactions on the Web*, 13(1):1:1–1:29, 2019.

[63] Yaxing Yao, Davide Lo Re, and Yang Wang. Folk models of online behavioral advertising. In *Proceedings of the Conference on Computer-Supported Cooperative Work and Social Computing (CSCW)*, pages 1957–1969, 2017.

# A Websites Analyzed

## Top Websites

adobe.com, aliexpress.com, amazon.com, ask.com, bbc.co.uk, bet9ja.com, booking.com, buzzfeed.com, cnn.com, coinmarketcap.com, craiglist.org, dailymail.co.uk, dailymotion.com, diply.com, discordapp.com, dropbox.com, ebay.com, etsy.com, facebook.com, github.com, google.com, indeed.com, mediafire.com, mozilla.org, nih.gov, nytimes.com, paypal.com, pinterest.com, providr.com, quora.com, reddit.com, roblox.com, rumble.com, salesforce.com, scribd.com, slideshare.net, spotify.com, stackexchange.com, stackoverflow.com, thestartmagazine.com, tumblr.com, twitch.tv, twitter.com, w3schools.com, whatsapp.com, wikia.com, wikihow.com, wikipedia.org, wordpress.com, yelp.com

### Middle Websites

17track.net, abcnews.go.com, avclub.com, babbel.com, bbb.org, cbc.ca, colorado.edu, desmos.com, file-upload.com, funsafetab.com, furaffinity.net, gamepress.gg, huawei.com, indiewire.com, intel.com, internshala.com, kijiji.ca, ladbible.com, mit.edu, myspace.com, news24.com, openclassrooms.com, opera.com, pathofexile.com, php.net, pixiv.net, poloniex.com, python.org, qwant.com, researchgate.net, rollingstone.com, runescape.com, sfgate.com, signupgenius.com, space.com, speedtest.net, theadvocate.com, trustedreviews.com, tufts.edu, ucl.ac.uk, umd.edu, ups.com, upsc.gov.in, utah.edu, wattpad.com, wikiwand.com, worldbank.org, worldoftanks.com, yifysubtitles.com, zapmeta.ws

### Bottom Websites

abebooks.com, adorama.com, artsy.net, bovada.lv, cj.com, classlink.com, coreldraw.com, dotloop.com, elitedaily.com, eurowings.com, fangraphs.com, filmapik.co, findlaw.com, fineartamerica.com, foodandwine.com, fronter.com, garena.com, gear4music.com, ghafla.com, hide.me, hsn.com, hsreplay.net, junkmail.co.za, justjared.com, kodi.tv, ldoceonline.com, letgo.com, lpu.in, majorgeeks.com, metacrawler.com, momjunction.com, mr-johal.com, ni.com, notepad-plus-plus.org, ou.edu, phys.org, playhearthstone.com, priceprice.com, rarlab.com, rice.edu, shein.in, statistic-showto.com, stocktwits.com, theathletic.com, tradingeconomics.com, uottawa.ca, uptostream.com, usgamer.net, volvocars.com, wimp.com

## B Website Analysis Template

**Step 1: Visit the homepage of the website**

1. Please enter the name of the website (use the format "google.com").
2. Did you see a notice for consumers that is an "opt-in" to the website's privacy policy and terms of conditions (including the use of cookies)? [Yes, and it included a way to opt-out or change settings; Yes, but it did not include a way opt-out or change settings; No]
3. Is there an option on the website to create a user account? [Yes, No, Other (please specify)]

   Logic: The following two questions are displayed if Q3 = Yes

**Step 2: Please create a user account for this site.**

4. Do you see the option to opt out of the site's marketing during the account creation process? [Yes, No, Other (please specify)]
5. Does the website have account settings? [Yes, No, Other (please specify)]

**Step 3: Look for an "about advertising" or "ad choices" related link on the home page. Click on the "about advertising" or "ad choices" link if it is there.**

6. Is there an "about advertising" or "ad choices" related link on the home page? [Yes, and it works; Yes, but it's broken; No]

   Logic: The following question is displayed if If Q6 = Yes, and it works or Q6 = Yes, but it's broken

7. What was this link labeled? [Ad Choices, Something else (copy label) ]

   Logic: The following three questions are displayed if Q6 = Yes, and it works

8. Where does the link direct you to? [Somewhere inside privacy policy, Somewhere inside account settings, An individual web page within the site that introduces OBA opt-outs, DAA's webpage, NAI's webpage, TrustE/TrustArc website, Other group's webpage]
9. By which parties are the advertising opt-outs on this page implemented? Include all entities that are linked to on the page. (select all that apply) [DAA, DAA of Canada (DAAC), European Interactive Digital Advertising Alliance (EDAA), Australian Digital Advertising Alliance (ADAA), NAI, TrustE/TrustArc service, The website, The browser or operating system (e.g., instructions to clear cookies or reset device advertising identifier), Google/Doubleclick, Other groups (please specify), There are no advertising opt-outs on this page]
10. How many user actions (e.g., clicks, form fields, hovers) are in the shortest path to completion out of all the opt-outs provided on this page?
11. What is the default setting for the opt-outs on this page (e.g., types of emails or ads already opted out of)? If none, enter 'NA'.

**Step 4: Now please go back to the homepage if you are not already there.**

12. Could you find the link to the site's privacy policy, or a page equivalent to a privacy policy? [Yes, and the link works; Yes, but the link is broken; No]

    Logic: The following six questions are displayed if Q12 = Yes, and the link works

**Step 5: Visit the website's privacy policy, or the page equivalent to a privacy policy. Some websites may call their privacy policy something else.**

13. Please copy and paste the URL for this page. Retrieve this policy through the policy retrieval tool.
14. Please copy and paste the title of the site's privacy policy.
15. Does the privacy policy (or equivalent page) have a table of contents? [Yes, No, Other (please specify)]

**Step 6.1: Next, do a search for "marketing," "e-mail," "email," "mailing," "subscribe," "communications," "preference" or "opt" in the privacy policy to look for marketing opt-outs. Also skim through the policy headings to double check.**

16. Does the privacy policy say that the site sends marketing or other types of communications (including email)? [Yes, the site sends communications, No, the site does not send communications, Not specified in the privacy policy, Other (please specify)]
17. Does the privacy policy have text about how to opt out of the site's marketing? [Yes, No, Not applicable (the site doesn't send marketing messages), Other (please specify)]

    Logic: The following six questions are displayed if Q16 = Yes

18. Please copy and paste the highest level heading in the policy where it describes how to opt out of the site's marketing.
19. Please copy and paste the paragraph(s) in the policy describing how to opt out of the site's marketing in the privacy policy.
20. According to the privacy policy, what types of communications can users opt out of receiving? (Make a note in the comment section if the first and third party emails are not clearly distinguished) [Newsletters, First-party marketing/promotional emails, Third-party marketing/promotional emails, User activity updates, Site announcements, Surveys, Mails, Phone calls, Text Messages/SMS, Other (please specify), None of the above]
21. According to the privacy policy, what types of communications users CANNOT opt out of? [Newsletters, First-party marketing/promotional emails, Third-party marketing/promotional emails, User activity updates, Site announcements, Surveys, Mails, Phone calls, Text Messages/SMS, Other (please specify), None of the above]
22. Does the privacy policy specify whether you can opt-out of marketing within the e-mails? [Yes, you can opt-out within the e-mails; Yes, but you can't opt-out with the e-mails; No, it wasn't specified]
23. Does the privacy policy include any links to marketing opt-outs? [Yes, there's one link to a marketing opt-out; Yes, there're multiple links to a marketing opt-out; No]

    Logic: The following four questions are displayed if Q23 = Yes, there's one link to a marketing opt-out or Q23 = Yes, there're multiple links to a marketing opt-out

**Step 6.2: Next, one by one click the links to the marketing opt-out links.**

24. Do any of the links in the privacy policy to the marketing opt-outs work? [Yes, they all work; Some work, but some do not; No, none of the links to the marketing opt-outs work]
25. Please copy and paste the URL(s) of the working links.
26. Please copy and paste the URL(s) of the broken links.
27. How many user actions (e.g., clicks, form fields, hovers) are in the shortest path to completion out of all the marketing opt-outs provided in the privacy policy?

    Logic: The following two questions are displayed if Q12 = Yes, and the link works

**Step 7.1: Next, do a search for "advertising," "ads," in the privacy policy in order to find whether the site has targeted advertising and their related opt-outs. Also skim through the policy headings to double check**

28. According to the privacy policy, does the website have targeted advertising? [Yes, the policy states there is targeted advertising; No, the policy states the website does not have targeted advertising; Not specified by the privacy policy]
29. Does the privacy policy page have text about how to opt out of the site's targeted advertising? [Yes, No, Not applicable (the site doesn't use OBA), Other (please specify)]

    Logic: The following seven questions are displayed if Q28 = Yes

30. Please copy and paste the highest level heading in the policy where it describes how to opt out of OBA.
31. Please copy and paste the paragraph(s) in the policy describing how to opt out of OBA.

32. According to the text of the privacy policy page, what can users opt out from related to OBA/tracking? [OBA only, Tracking, Not specified, Other (please specify)]

33. Does the privacy policy page say whether the OBA opt-outs located in the privacy policy will be effective across different browsers? [Yes, the policy says they will be effective across different browsers; Yes, but the policy says there're for current browser only; Not specified by the privacy policy; Other (please specify)]

34. Does the privacy policy page say whether the OBA opt-outs located in the privacy policy will be effective across different devices? [Yes, the policy says they will be effective across different device; Yes, but the policy says there're for current device only; Not specified by the privacy policy; Other (please specify)]

35. By which parties are the OBA opt-outs mentioned by the privacy policy implemented? Include all entities that are linked to from the privacy policy. [DAA, DAA of Canada (DAAC), European Interactive Digital Advertising Alliance (EDAA), Australian Digital Advertising Alliance (ADAA), NAI, TrustE/TrustArc service, The website, The browser or operating system (e.g., instructions to clear cookies or reset device advertising identifier), Google/Doubleclick, Other groups (please specify)]

36. Does the privacy policy page include any links to an OBA opt-out? [Yes, there is one link to an OBA opt-out; Yes, there're multiple links to different OBA opt-outs; Yes, there're multiple links to same OBA opt-out; No]

    Logic: The following four questions are displayed if Q35 = Yes, there is one link to an OBA opt-out or Q35 = Yes, there're multiple links to different OBA opt-out

    **Step 7.2: Next, one by one click the links to the OBA opt-outs in the privacy policy.**

37. Do any of the links in the privacy policy to the OBA opt-outs work? Note: Count links with different text and the same URL as multiple links. Include links from the privacy policy and one layer of linked pages as well. [Yes, they all work; Some work, but some do not; No, none of the OBA opt-out links work]

38. Please copy and paste the URL(s) of the working links. Place each URL on its own line.

39. Please copy and paste the URL(s) of the broken links. Place each URL on its own line.

40. How many user actions (e.g., clicks, form fields, hovers) are in the shortest path to completion out of all the OBA opt-outs provided in the privacy policy?

41. What is the default setting for the OBA opt-outs in the privacy policy (e.g., types of emails or ads already opted out of)? If none, enter 'NA'.

    Logic: The following question is displayed if Q12 = Yes, and the link works

    **Step 8.1: Next, do a search for "delete," "deletion,""closing account," "remove" or similar terms in the privacy policy in order to find data deletion choices. Also skim through the policy headings to double check.**

42. Is there any information in the privacy policy that introduces how to delete your account data? [Yes, No, Other (please specify)]

    Logic: The following eight questions is displayed if Q42 = Yes

43. Please copy and paste the highest level heading in the policy where it describes how to delete account data.

44. Please copy and paste the paragraph(s) in the policy where it describes how to delete account data.

45. According to the privacy policy, what actions can users perform related to data deletion? [Delete their account permanently, Suspend/deactivate their account (data will not be permanently deleted right away), Choose specific types of data to be deleted from their account, Not specified, Other (please specify)]

46. Please copy and paste the specific types of data indicated in the privacy policy.

47. According to the privacy policy, does the website suspend or deactivate your account before deleting it? [Yes, the policy says your account will be suspended; No, the policy says your account will be deleted after a certain amount of time; Not specified in the policy; Other (please specify)]

48. According to the privacy policy, after how long will the data be permanently deleted? [Not specified, Immediately, One week, 30 days, 60 days, 90 days, 6 months, Other (please specify)]

49. How many user actions (e.g., clicks, form fields, hovers) are in the shortest path to completion out of all the data deletion options?

50. Does the privacy policy include any links to delete your account data? [Yes, there's one link; Yes, there're multiple links; No]

    Logic: The following three questions are displayed if Q50 = Yes, there're one link or Q50 = Yes, there're multiple links

**Step 8.2: Next, one by one click the links to the data deletion choices.**

51. Does the link in the privacy policy to the data deletion choice work? [Yes, they all work; Some work, but some do not; No, they're all broken]

52. Please copy and paste the URL(s) of the working links.

53. Please copy and paste the URL(s) of the broken links.

Logic: The following five questions are displayed if Q11 = Yes, and the link works

**Step 9: Next, search for "Do Not Track" or "DNT" in the privacy policy.**

54. Will the website honor DNT requests? [Yes, No, Not specified in the privacy policy]

**Step 10: Next, skim through the policy for things users can opt-out of. Adjust your previous answers if necessary and complete the following questions.**

55. Did you find any other type of opt-outs in the privacy policy? [Yes, No]

56. What other things can users opt out from at this site as described in the privacy policy? [Device info; All first-party cookies; Location history; Profile activities/inferred interests; Sharing with third parties; Google Analytics; Other (please specify); None of the above]

57. When you are skimming through the privacy policy, could you find any other pages that aim to explain the privacy policy or the privacy and data practices of the company in general? [Yes, and the link works; Yes, but the link is broken; No; Other (please specify)]

58. Please copy and paste the URL of the link(s).

59. Did the privacy policy describe the location of a marketing or communications opt out located in the account settings? [Yes, No]

**Step 11: Go to this described location in the account settings or look through the main levels of the account settings for marketing, email, or communication choices. Click links which seem to indicate user choice or preferences.**

60. Is there any marketing opt-out located in the account settings? [Yes, No, Not applicable (the site doesn't send email/marketing messages), Other (please specify)]

61. How many user actions (e.g., clicks, form fields, hovers) are in the shortest path to completion to this marketing opt-out?

62. What is the default setting for the marketing opt-outs in the account settings (e.g., types of emails or ads already opted out of)? If none, enter 'NA'."

63. Is it the same marketing opt-out page that was presented in the privacy policy? [Yes; No, it's a different marketing opt-out page; There was no marketing opt-out described in the privacy policy; Other (please specify)]

Logic: The following question is displayed if Q63 is not "Yes"

64. What types of communications can users opt out of from in the account settings? [Newsletters, First-party marketing/promotional emails, Third-party marketing/promotional emails, User activity updates, Site announcements, Surveys, Mails, Phone calls, Text Messages/SMS, Other (please specify), None of the above]

65. Did the privacy policy describe the location of an OBA opt-out located in the account settings? [Yes, No]

**Step 12: Go to this described location in the account settings or look through the main levels of the account settings for advertising choices. Click links which seem to indicate user choice or preferences.**

66. Is there any OBA opt-out located in the account settings? [Yes, No, Not applicable (the site doesn't use OBA), Other (please specify)]

67. How many user actions (e.g., clicks, form fields, hovers) are in the shortest path to completion to this targeted advertising opt-out?

68. Is it the same opt-out page that was presented in the privacy policy? [Yes; No, it's a different OBA opt-out page; There was no OBA opt-out described in the privacy policy; Other (please specify)]

Logic: The following four questions are displayed if Q68 is not "Yes"

69. By which parties is the OBA opt-out in the account settings implemented? Include all entities that are linked to from the account settings. [DAA, DAA of Canada (DAAC), European Interactive Digital Advertising Alliance (EDAA), Australian Digital Advertising Alliance (ADAA), NAI, TrustE/TrustArc service, The website, The browser or operating system (e.g., instructions to clear cookies or reset device advertising identifier), Google/Doubleclick, Other groups (please specify)]

70. What can users opt out from related to OBA/tracking from the account settings? [OBA only (users will still be tracked), Tracking, Not specified, Other (please specify)]

71. According to the information provided, will the OBA opt-out in the account settings be effective across different browsers? [Yes; No, it's for current browser only; Not specified; Other (please specify)]

72. According to the information provided, will the OBA opt-out in the account settings be effective across different devices? [Yes; No, it's for current device only; Not specified; Other (please specify)]

73. Did the privacy policy describe the location of a data deletion choice in the account settings? [Yes, No]

**Step 13: Go to this described location in the account settings or look through the main levels of the account settings for data deletion choices. Click links which seem to indicate user choice or preferences.**

74. Is there any data deletion option located in the account settings? [Yes, No, Other (please specify)]

75. How many user actions (e.g., clicks, form fields, hovers) are in the shortest path to completion to this data deletion option?

76. Is it the same data deletion page that was presented in the privacy policy? [Yes; No, it's a different data deletion page; There was no data deletion choice presented in the privacy policy; Other (please specify)]

Logic: The following four questions are displayed if Q76 is not "Yes"

**Step 14: Lastly, look through the main levels of the account settings for other types of user choices. Click links which seem to indicate user choice or preferences.**

81. Did you find any other opt-outs in the account settings? [Yes, No]

77. According to the information provided, what actions can users perform related to data deletion? [Delete their account permanently, Suspend/deactivate their account (data will not be permanently deleted right away), Choose specific types of data to be deleted from their account, Not specified, Other (please specify)]

78. Please copy and paste the specific types of data it indicates. Use ";" to separate multiple items.

79. According to the information provided, does the website suspend or deactivate your account before deleting it? [Yes, there's information that says your account will be suspended; No, there's information that says your account will be deleted after a certain amount of time; Not specified within the account settings; Other (please specify)]

80. According to the privacy policy, after how long will the data be permanently deleted? [Not specified, Immediately, One week, 30 days, 60 days, 90 days, 6 months, Other (please specify)]

82. What other things can users opt out from in the account settings? [Device info; All first-party cookies; Location history; Profile activities/inferred interests; Sharing with third parties; Google Analytics; Other (please specify); None of the above]

# The Fog of Warnings: How Non-essential Notifications Blur with Security Warnings

Anthony Vance, *Temple University* David Eargle, *University of Colorado Boulder*
Jeffrey L. Jenkins, *Brigham Young University* C. Brock Kirwan, *Brigham Young University*
Bonnie Brinton Anderson, *Brigham Young University*

## Abstract

Adherence to security warnings continues to be an important problem in information security. Although users may fail to heed a security warning for a variety of reasons, a major contributor is habituation, which is decreased response to repeated stimulation. However, the scope of this problem may actually be much broader than previously thought because of the neurobiological phenomenon of generalization. Whereas habituation describes a diminished response with repetitions of the same stimulus, generalization occurs when habituation to one stimulus carries over to other novel stimuli that are similar in appearance.

Generalization has important implications for the domains of usable security and human–computer interaction. Because a basic principle of user interface design is visual consistency, generalization suggests that through exposure to frequent non-security-related notifications (e.g., dialogs, alerts, confirmations, etc.) that share a similar look and feel, users may become deeply habituated to critical security warnings that they have never seen before. Further, with the increasing number of notifications in our lives across a range of mobile, Internet of Things, and computing devices, the accumulated effect of generalization may be substantial. However, this problem has not been empirically examined before.

This paper contributes by measuring the impacts of generalization in terms of (1) diminished attention via mouse cursor tracking and (2) users' ability to behaviorally adhere to security warnings. Through an online experiment, we find that:

- Habituation to a frequent non-security-related notification does carry over to a one-time security warning.
- Generalization of habituation is manifest both in (1) decreased attention to warnings and (2) lower warning adherence behavior.
- The carry-over effect, most importantly, is due to generalization, and not fatigue.
- The degree that generalization occurs depends on the similarity in look and feel between a notification and warning.

These findings open new avenues of research and provide guidance to software developers for creating warnings that are more resistant to the effects of generalization of habituation, thereby improving users' security warning adherence.

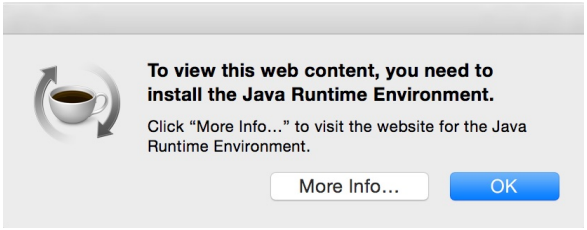


*USENIX Symposium on Usable Privacy and Security (SOUPS) 2019*.
August 11 -- 13, 2019, Santa Clara, CA, USA.

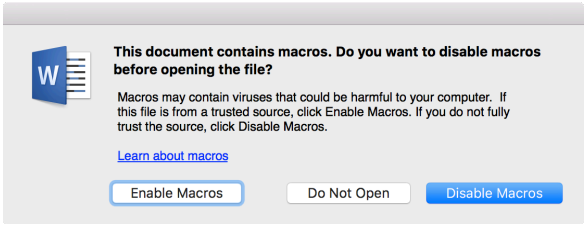
## 1. Introduction

Users' adherence to security warnings continues to be an important problem in information security because warnings are often the last defense standing between a user and compromise [1, 36]. Although users may fail to heed a warning for a variety of reasons [24], an important contributor is habituation, which is defined as decreased response to repeated stimulation [9, 11, 18, 19, 25]. This phenomenon is fundamentally neurobiological in nature [23], and past work has shown how the brain habituates to security warnings over time [5, 34].

However, there is a key aspect of neurobiology's habituation theory that has not been examined but that has critical implications for security warnings. *Stimulus generalization*—or simply *generalization*—occurs when the effects of habituation to one stimulus *generalize*, or carry over, to other novel stimuli that are similar in appearance [23, 31]. Applied to the domain of human–computer interaction, generalization suggests that users not only habituate to individual security warnings, but also to whole classes of user interface (UI) notifications (e.g., dialogs, alerts, confirmations, etc.—hereafter referred to collectively as "notifications" for brevity) that share a similar look and feel (see Figure 1).




System-generated notification




Security warning

**Figure 1: A notification and security warning. Note the similarities in UI and mode of interaction.**

Consistency of look and feel is a foundational principle in UI design [14, 21] and is reinforced by major software companies, such as Apple and Microsoft, which provide development libraries and guidelines to ensure consistency across software applications [8, 22]. As a result, users may already be deeply habituated to a security warning that they have never seen before.

With the increasing number of notifications in the lives of users across a range of mobile, Internet of Things, and computing devices, the accumulated effect of generalization may be substantial, lessening the effectiveness of comparatively rare security warnings that are truly critical. For example, an analysis of 40,191 Android users showed that they received an average of 26 notifications per day on their mobile devices, not including apps that "flood" users with notifications, such as Skype, Viber, and DropSync [26]. In such a saturated environment, it is crucial that habituation to notifications not generalize to security warnings; the latter are to have protective value.

Although the problem of the blurring of security warnings and notifications has previously been recognized (e.g., [9, 33]), it has not been empirically studied. Consequently, the scope and severity of generalization, as well as the conditions under which it occurs, are not known. By measuring these things, we can better understand how generalization occurs and mitigate its influence.

The objective of this research is to measure and explain how habituation to a frequent non-security-related notification generalizes or carries over to security warnings. In doing so, we answer the following research questions:

> RQ1: Does habituation to non-security-related notifications generalize to security warnings?
>
> RQ2: Does the degree of look-and-feel similarity influence the amount of generalization of habituation?

Using mouse cursor tracking and other behavioral responses in an online experiment, we show that:

- Habituation to a frequent non-security-related notification does carry over to a one-time security warning.
- Generalization of habituation is manifest both in (1) decreased attention to warnings and (2) lower warning adherence behavior.
- Importantly, we show that this carry-over effect is due to generalization, and not fatigue.
- The degree that generalization occurs depends on the similarity in look and feel between a notification and warning.

These findings help form a foundation for developing warning designs that are resistant to the influence of generalization.

## 2. Related Work

### Generalization in Useable Security Research

Although habituation to security warnings is well known and has been examined in a number of studies [10-12, 38], the phenomenon of generalization is less well recognized. West noted that "security messages often resemble other message dialogs. As a result, security messages may not stand out in importance and users often learn to disregard them" [37, p. 39]. Böhme and Köpsell observed that a user's automatic response to notifications "seems to spill over from moderately relevant topics (e.g., EULAs) to more critical ones (online safety and privacy)" [9, p. 2406]. However, neither of these studies empirically examined this effect.

Similarly, researchers have observed that habituation to a single warning in one context can carry over to a different context. For example, Egelman et al. [15] observed that some lab participants disregarded a phishing warning because they confused it with a previous warning they had seen. However, this was an incidental observation and not the focus of their study. They speculated that warning visual similarity caused the confusion, but they did not test this supposition. Similarly, Sunshine et al. [29] observed that users who correctly identified the risks of an SSL warning in a library context inappropriately identified these same risks in a banking context. Likewise, Amer et al. [3] found that users who habituated to exception notifications in one context were habituated to a different though visually identical exception notification in a different context. However, in each of these cases, the users habituated to the same type of security warning or notification. As a result, it is unclear to what extent software notifications generalize to security warnings.

### Generalization in Neuroscience Research

As users respond repeatedly to notifications, they are likely to devote fewer neural resources toward those stimuli, either through habituation or through perceptual learning [4, 6, 7, 34]. Perceptual learning occurs when there is a structural change in visual processing structures of the brain to support performance on a perceptual task as a result of previous visual experience [16]. The neuroscience literature has long shown that this increased efficiency of the neural response comes at the price of generalizing from one stimulus set to another similar set of stimuli [31].

Generalization has been demonstrated in the neuroscience literature at a number of different levels [31], including decreased neural responses to stimuli similar to habituated stimuli [23], the transfer of perceptual learning to novel tasks [13], and the retrieval of long-term memory representations to similar memory cues [20]. Habituation is typically short-lived, as neural responses typically return to baseline after a delay. Conversely, perceptual learning can be long-lasting, can occur without overt attention [13], and is more likely to be involved in more complex tasks (such as using complex software) [17].

## 3. Methods

In order to examine generalization, we designed an online experiment to measure habituation (a prerequisite condition for generalization), generalization, and warning adherence behavior. Research shows that people are not very accurate in self-reporting security behavior [35], so we instead captured direct behavioral measures. First, we measured habituation in terms of the mousing speed of users' responses to notifications and warnings as measured via mouse cursor tracking. Previous research has demonstrated this to be a robust measure of habituation to security warnings [5, 33, 34]. Similarly, we also measured habituation in terms of the time between the display of a notification or warning and when a user responded to it. Finally, we also measured users' adherence to the security warning, "the rates at which users do not proceed through a warning, i.e., the rate at which they choose the safer option" [24, p.7].

### Participants

We recruited 600 participants via Amazon Mechanical Turk (mTurk). Following Steelman et al. [28], all participants were required to be from the United States. The average age of participants was 36 years old (min: 18, max 76); 53% were male. Participants were ultimately paid $1.50 ($1.00 up front, with a $0.50 bonus) for an approximately five-minute task. Table 1 shows the participant breakdown per condition.

### Ethics

The university Institutional Review Board (IRB) approved the protocol used. In an informed consent statement, participants were told that the study objective was to determine how people visually evaluate and cognitively process computer software messages. They were also told that in the experimental task they would be browsing websites and perform simple tasks such as comparing images. However, participants were not told that we were specifically interested in their response to security warnings. At the end of the experiment, participants were debriefed about the specific objectives of the experiment.

### Experimental Task

We followed a previously established experimental protocol in which participants classified images on the web as either animated or photographic versions of Batman [32]. Participants from Amazon Mechanical Turk were required to use the Firefox browser and were directed to a server on which we hosted our experiment. A dashboard allowed participants to classify each loaded image (Figure 2).

In pre-task instructions, participants were told that random webpages containing images of Batman would be loaded into a central frame on the task dashboard. Using the following language, participants were told that because the sites that would be loaded were random and external, some risk to their devices was involved:

"Warning: The researchers are not responsible for the content of the webpages loaded into the center frame. By participating in this task, you understand that despite the pages being in a center frame, the risks are the same as if you were visiting the pages directly. You assume all risks associated with visiting these websites."

Participants went through a task warm-up "internet connectivity" test where two actual live external pages were loaded into the central frame, which participants were instructed to interact with and peruse. However, in reality, the main Batman classification task loaded *static* screenshot images of websites with photos of Batman into the central dashboard frame. This allowed us to control what participants saw during the task.

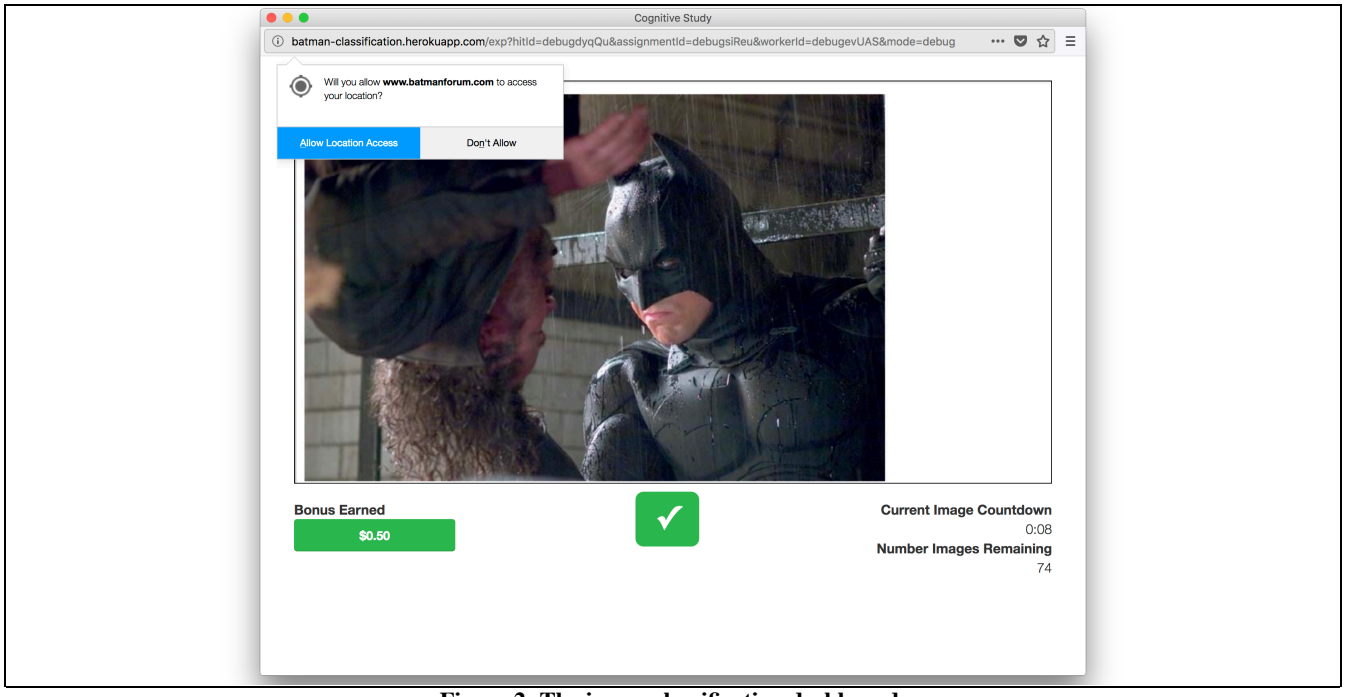



**Figure 2: The image classification dashboard.**

We reasoned that if participants thought that the task was loading real external websites, then they would be more likely to believe that the appearance of a popup security notification was triggered by the loaded external Batman website, as opposed to by the experiment dashboard. The source URLs that we put int the text of some of the security warnings reinforced the perception that the external sites were triggering the warnings to appear. We also encouraged a belief that the task loaded unregulated external websites in a bid to dampen the likelihood of lab experiment bias [27], wherein participants may feel an invincibility against threats because they feel secure within the walled confines of an artificial experiment approved by an ethics board. Our analysis suggested that participants believed security popups were real (see section 4.1).

Participants were under time pressure to complete the task. For each website, participants had ten seconds in which to classify the image. Failure to classify the image was counted as an incorrect answer. A performance bar in the bottom-left corner of the screen provided participants with live feedback of their current bonus standing. Initially, the bar was green, but an incorrect classification decreased a participant's bonus by 5 cents, updating the bonus bar with a depressing red slider animation from the right side to represent the loss. We had the bonus be dependent on performance in order to encourage continued participant engagement with the task. In reality, however, all participants received the full bonus regardless of their performance. They were informed of their full reward as part of the post-task debrief.

After the internet connectivity test and instructions, participants first completed a warm-up round of four Batman image classifications, during which no popups or security warnings appeared, before beginning what they thought would be 75 total image classifications. After each classification in the non-warmup 75-set, a HTML5-styled notification styled after the Firefox location permission request reported the participant's current classification performance (see Figure 3). Importantly, participants had to click a "continue" button on this performance notification before going on to the next image, thus forcing them to interact with each notification. Each participants encountered a single randomly-assigned security warning during their task after a randomly-assigned number of interactions with Batman image classifications and performance notifications. Once participants saw their security warning, the main classification experimental task abruptly terminated. Javascript recorded all participant interactions during the task, including mouse cursor movements, reaction times, and security warning choice click-behavior. Following the main task, participants were directed to a short post-task survey and debrief, after which the experiment was complete.

In summary, we chose the Batman protocol because it provided an excuse to show participants, who were using their own computers, multiple ostensibly-real browser task-related notifications within a short timeframe, one of which was a security notification supposedly triggered by a non-experimenter-controlled external website, in a closely-web-observable (through javascript) environment.

### Experimental Treatments

To answer our research questions, we randomly assigned participants to 1 of 10 experimental conditions in a 2

(manipulating generalization) × 5 (manipulating the similarity of the look-and-feel) factorial experimental design (Table 1). First, we manipulate generalization by either displaying the warning first or after a series of notifications. Second, we manipulate how similar the look-and-feel is between the notification and the target stimulus (using four security warnings in Firefox with varying look-and-feel similarity to the notification, and a novel stimulus). We describe these manipulations in more detail below.

**Table 1: Experimental Design (2x5, fully-crossed) with cell *n*'s.**

| | **Appeared After Classification** | |
|---|---|---|
| **Security Warning Type** | Position 1 | Position 15 |
| Permission warning | $n = 59$ | $n = 60$ |
| Extension warning | $n = 60$ | $n = 61$ |
| Save executable | $n = 60$ | $n = 60$ |
| Open macro | $n = 60$ | $n = 60$ |
| Novel stimulus | $n = 60$ | $n = 60$ |

In order to assess whether habituation to notifications generalized to security warnings, we first manipulated whether participants were habituated to notifications by assigning them to view a security warning either after the *first* Batman image classification or after the *fifteenth* image classification. By measuring responses to warnings at both positions, we could measure and control for differences within each security warning type between its two appearance positions, as well as calculate differences across security warning types for a given position. Participants who were in one the "position 15" treatment groups classified 15 Batman images, with a performance notification being shown after each of the first 14 Batman classifications, and their assigned security warning being shown after the 15th Batman image classification instead of a performance notification, followed by an abrupt task termination. Participants who were in one of the "position 1" treatment groups only classified one Batman image, after which they saw and interacted with their assigned security warning, followed by an abrupt task termination. This means that participants who saw a security warning position 1 did not see *any* performance notifications – they only saw one Batman and one security notification.

We also manipulated the type and look of the security warning. Participants were randomly assigned to view either one of four different simulated Firefox security warnings, or a visually novel stimulus (described in section 3.5). The Firefox security warnings were chosen because they had varying levels of look-and-feel similarity to the task-performance notification, which helps address our second research question (see Figure 3). The most visually similar security warning to the performance notification was the location permission warning ("permission warning"; Figure 4); the second-most visually-similar was a Firefox add-on installation permission warning ("extension warning";

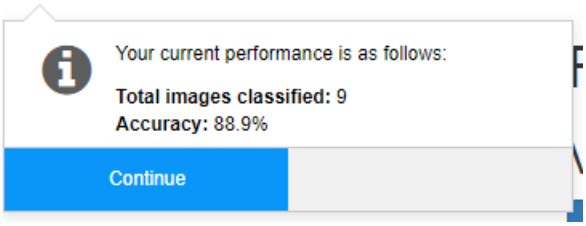



**Figure 3: HTML5 performance notification.**

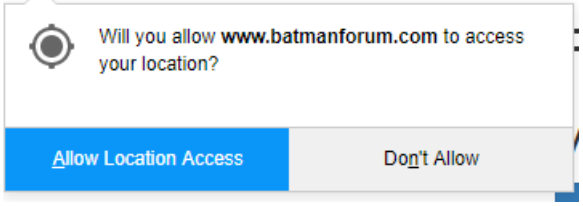



**Figure 4: HTML5 permission warning.**

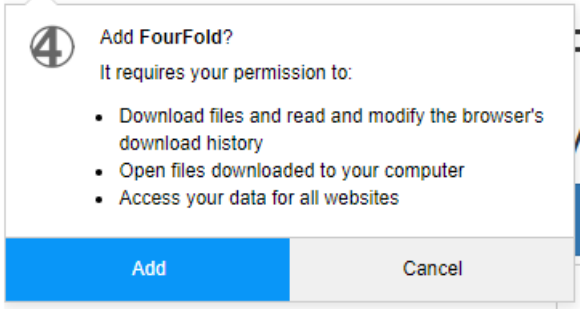



**Figure 5: Firefox add-on (extension) warning.**

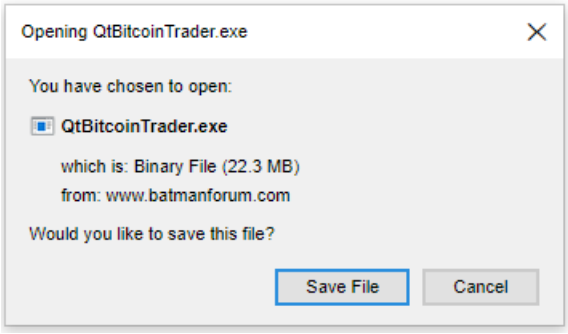



**Figure 6: Firefox save executable message.**

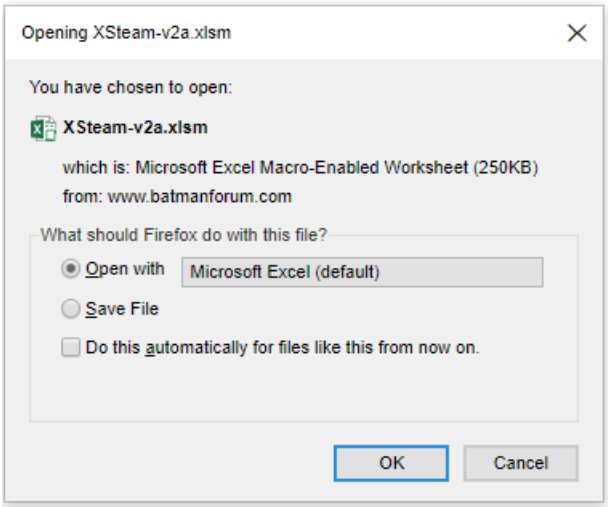



**Figure 7: Firefox open macro-enabled spreadsheet message.**

Figure 5); and the most visually discrepant security warnings compared to the performance notification were Firefox save-executable message ("executable save"; Figure 6), and a Firefox 'open a macro-enabled spreadsheet' message ("open macro"; Figure 7). Each of these four fake security notifications were designed in HTML5 and javascript to look just as would their legitimate Firefox warning counterparts.

We recognize that the save executable and open macro messages are not security warnings, strictly speaking, because they do not actually warn the user of anything. However, these messages do have strong security implications. In particular, opening documents with malicious macros are a longtime and increasingly popular avenue of attack [30]. For simplicity, we refer to all of our security message treatments as warnings.

**Ruling out the effect of fatigue**

To rule out the effect of fatigue, we designed a treatment that was visually novel compared to the other notification and security warnings (Figure 8). Following the neurobiological literature [23], generalization of habituation is measured by showing that once a participant habituates to a stimulus, a neural or behavioral response shows little increase when a novel stimulus is presented that is similar to the original stimulus. However, when a novel stimulus—an image of a yellow duck—is presented that is very different from the original stimulus, the response recovers to where it was before any stimuli were displayed, thus demonstrating that fatigue was not the reason for the diminished response to similar stimuli [23]. Participants assigned to the novel-stimulus condition saw it at either position 1 and position 15, which allowed us to test for differences between positions. Any slower reaction times between participants who saw the duck at position 15 versus position 1 would be indicative of fatigue or of general task dismiss-the-notification familiarity for the former group. If there was evidence of such fatigue within the duck position-treatments, then we could control for that magnitude of fatigue in our other security warning tests.

## 4. Analysis

**Realism check**

The real-website ruse worked—participants were successfully led to believe that security warnings were triggered by the loaded websites they automatically visited. Both quantitative and qualitative (after the debrief) responses from participants supported that they held this belief. For instances, in a free-response field on the post-task survey, one participant said "The pop up was unexpected and I thought I might have clicked on something wrong. I did pause for a second and panic," and another said "That was incredible deception. I am a software engineer with a background in cybersecurity and you fooled [me]." A third stated, "I got bamboozled." When asked in the survey about their perceived realism of the security messages that they

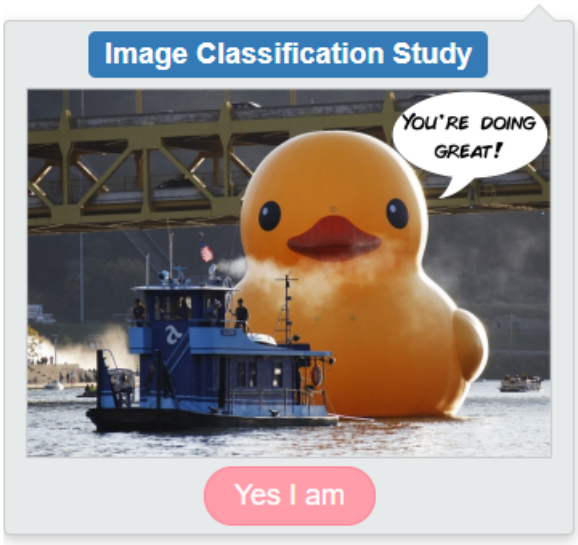



**Figure 8: Novel stimulus for assessing fatigue.**

saw, participants rated the security message mockups well above 5 out of 10 (see Figure 9).

**Adherence Behavior**

We measured whether participants who saw a security warning clicked through it (e.g., taking the "accept" or "proceed" action for one of Figures 5–8). By comparing click-through rates for each security warning between its two appearance positions, we can test whether generalization had an impact on an actual security behavior — which it indeed did.

We built a logistic regression model including only those who received warnings and not the novel stimulus ($N = 487$, Nagelkerke's $R^2 = .546$), which predicted whether participants clicked through their security warning. A click-through was coded as a 1, and any action that dismissed the warning without clicking through was coded as a 0. Independent variables were the security warning type (permission warning, extension warning, save executable,

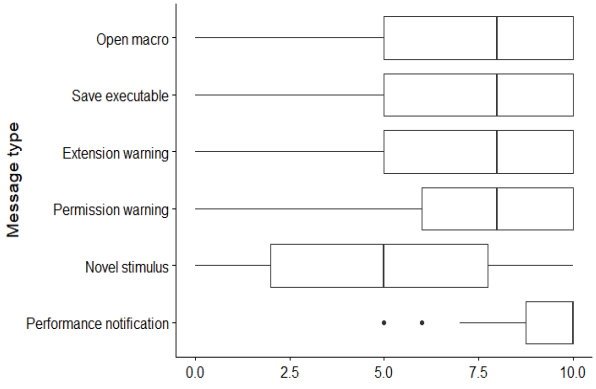



**Figure 9. Realism of message (self-reported, scale of 0 to 10). Note that perceived realism was not required for the novel stimulus.**

open macro), crossed with the position or order in which the warning was displayed (position 1, position 15). The model fit is shown in Table 2

The permission warning was more likely to be clicked through if seen at position 15 than at position 1 (OR = 2.60, $p = 0.008$), as was the extension request (OR = 1.95, *one-tailed* $p = .047$). No differences in click-through behaviors between positions were observed for either the open-macro (OR=0.59, $p = .192$) or the save-executable warnings (logOdds = 1.00, $p = 1.00$) (see Figure 10 and Table 2). As the permission request and extension request are more

visually similar to the performance notification than the open-macro and save-executable warnings, these findings support that the similar look-and-feel of security warnings to other notifications may be magnifying generalization.

**Mouse cursor movement speed**

As an indicator of habituation, we used mouse cursor movement speed as a dependent variable to test whether habituation to non-security notifications generalizes to security warnings. Movement speed refers to how fast a user moves over the warning to dismiss or adhere to the warning (in pixels traversed per millisecond). Faster movement speed indicates that the user is paying less attention to the content of the warning, and that the user is providing a habituated response to the warning. Slower movement speed indicates that the user is paying more attention to the warning and providing a non-habituated response to the warning [33].

**Table 2. Click-through predicted by interaction of warning type and appearance position, 0-intercept for ease of interpreting within-type slopes.**

did_click_through ~ 0 + security_message + security_message:showSecurityMessageAt

| | **Clicked-through** | | |
|---|---|---|---|
| *Predictors* | *Odds Ratios (OR)* | *CI* | *P (one-tailed)* |
| Permission warning | 0.33 | 0.18 – 0.58 | **<0.001** |
| Extension warning | 0.33 | 0.19 – 0.60 | **<0.001** |
| Save-Executable warning | 0.03 | 0.01 – 0.14 | **<0.001** |
| Open-macro warning | 0.15 | 0.07 – 0.32 | **<0.001** |
| Permission warning × position 15 | 2.60 | 1.20 – 5.62 | **0.008** |
| Extension warning × position 15 | 1.95 | 0.89 – 4.24 | **0.047** |
| Save-executable × position 15 | 1.00 | 0.14 – 7.34 | 1.000 |
| Open-macro × position 15 | 0.59 | 0.18 – 1.92 | 0.192 |
| Observations | 487 | | |
| Cox & Snell's $R^2$ / Nagelkerke's $R^2$ | 0.409 / 0.546 | | |

We conducted several analyses to examine how generalization influences movement speed. First, we limited the data just to the warnings, and examined whether the position of the warning (1 or 15) influences movement speed. If the position of the warning influences movement speed, this indicates that habituation to the non-security

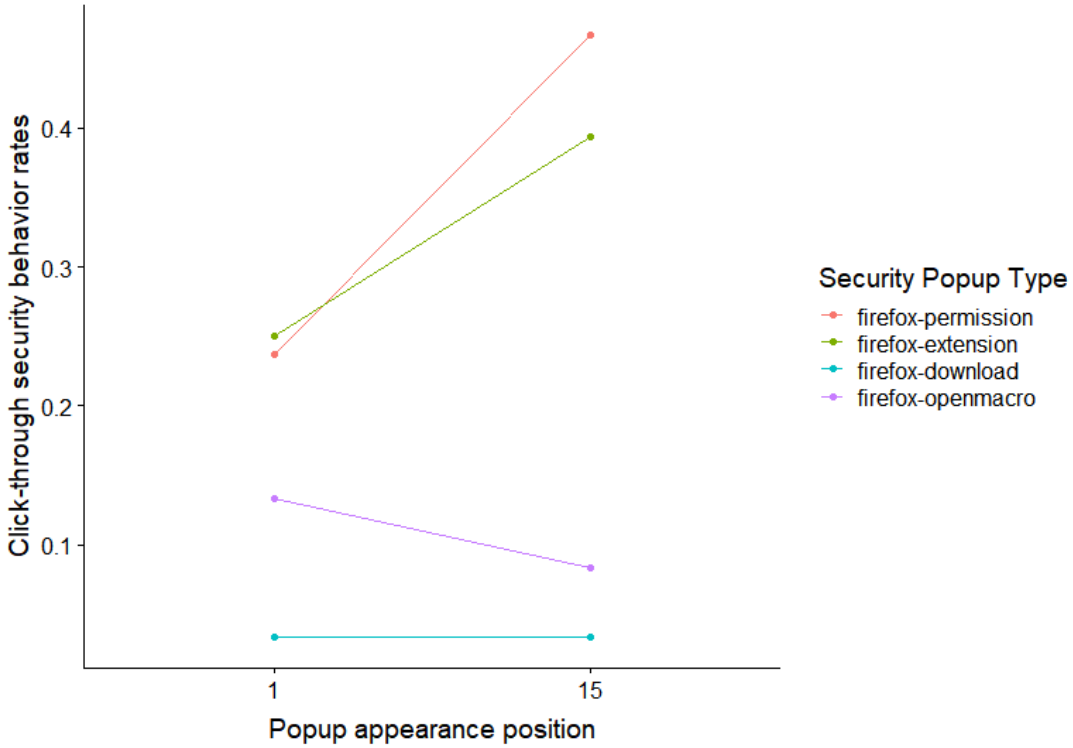



**Figure 10: Adherence behavior at positions 1 and 15.**

notifications is generalizing to the security notifications. Otherwise, there should be no significant difference. We specified a linear mixed model predicting movement speed by position. The type of warning was treated as a random effect. Position was treated as a fixed effect and was coded as 0 if the security notification was first, or 1 if the security warning occurred in position fifteen. The position significantly predicted speed: $t(449.004) = 5.471$, $p < .001$, conditional $R^2$: 0.231, supporting that generalization occurs (see Table 3).

To help ensure that the differences observed are due to generalization and not to fatigue, we specified a general linear model examining the influence of position on movement speed for the novel stimulus. In this analysis, position (1 vs. 15) did not influence how fast someone responded to the notification (see Table 4). This suggests that generalization rather than fatigue influenced movement speed.

**Table 3: Mixed linear regression predicting speed (px/ms) by position.**

| | Estimate | Std. Error | *df* | *t*-value | *Pr*(>\|*t*\|) |
|---|---|---|---|---|---|
| (Intercept) | 0.482 | 0.087 | 3.213 | 5.551 | 0.010 |
| position | 0.178 | 0.032 | 449.004 | 5.471 | < 0.001 |

**Table 4: Linear regression predicting speed (px/ms) based on position for novel stimulus.**

| | Estimate | Std. Error | *t*-value | *Pr*(>\|*t*\|) |
|---|---|---|---|---|
| (Intercept) | 0.574 | 0.050 | 11.535 | < 0.001 |
| position | -0.100 | 0.071 | -1.412 | 0.161 |

Finally, we examined whether the type of notification influenced the amount of generalization. To do this, we conducted a general linear model examining the interactions between the security warning types and position. Each warning type was coded as a dummy variable, leaving the performance notification as the baseline condition. Again, order was coded as a 0 if the notification was the first one shown. Otherwise, it was coded as a 1 if it was the fifteenth notification shown. The results are shown in Table 5. Although the main effects of warning type were significant, only the interactions (slope modifiers) for the extension warning and the permission warning with order were significant. These two types of warnings generalized less when compared to the non-security notification. The trends in speed for each notification type are shown in Figure 11. Again, the permission request and extension request are more visually similar to the performance notification than the macro and save executable warnings, these findings support that the similar look-and-feel of security warnings to other notifications may be magnifying generalization.

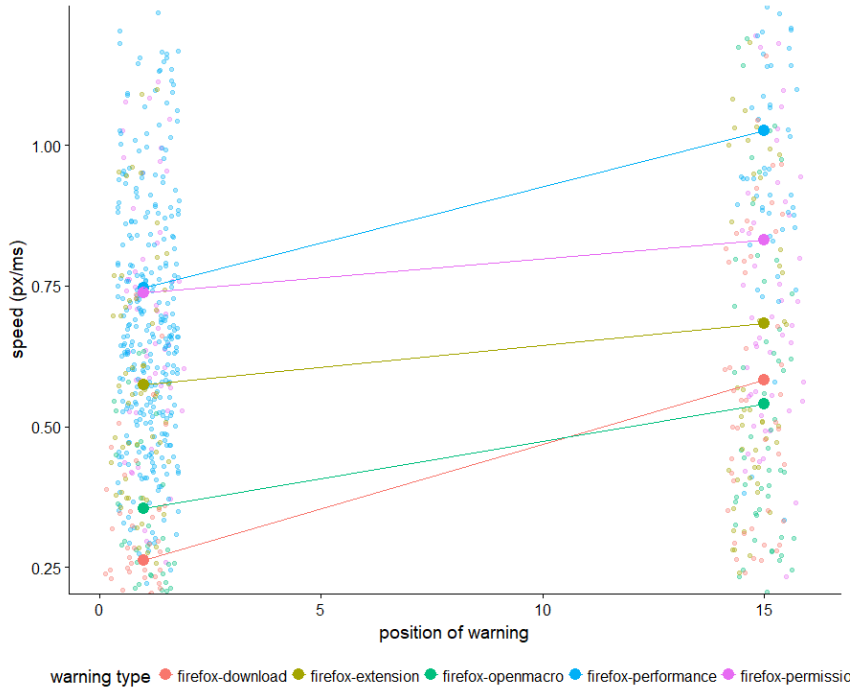



**Figure 11: Movement speed (px/ms) trend for each warning / notification type.**

**Table 5: Linear regression predicting speed (pixel per millisecond) by interaction of security warning type by appearance position.**

| | Estimate | Std.Error | t-value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.727 | 0.020 | 36.795 | 0.000 |
| Position | 0.020 | 0.003 | 5.824 | 0.000 |
| Extension warning | (0.161) | 0.052 | (3.097) | 0.002 |
| Save executable | (0.488) | 0.051 | (9.479) | 0.000 |
| Open macro | (0.387) | 0.051 | (7.511) | 0.000 |
| Permission warning | 0.004 | 0.051 | 0.070 | 0.944 |
| **Position × extension** | **(0.012)** | **0.006** | **(2.128)** | **0.034** |
| Position × executable | 0.003 | 0.006 | 0.531 | 0.595 |
| Position × open macro | (0.007) | 0.006 | (1.174) | 0.241 |
| **Position × permission** | **(0.013)** | **0.006** | **(2.338)** | **0.020** |

**Reaction Times**

We induced a linear model to examine the impact of warning type and appearance position on user reaction times. All reaction times greater than 2.5 standard deviations from the median (median = 1,447 ms, SD = 2,732 ms) were flagged as outliers and were summarily ousted. The remaining reaction times were subjected to a linear regression model, wherein they were predicted by the interaction of modal position and modal types (dummy-coded) (see Table 6 and Figure 12). The slope for the novel stimulus between positions one and fifteen was not significantly different from 0 ($\beta = -13.2, SE = 16.76, t = -0.79, p = 0.431$). This supports the notion that fatigue was not at play over the course of the experimental task.

The slope for the performance notification was precipitous (see Figure 12), flattening out around four exposures, as would be expected given that this warning appeared often in the classification task. Interestingly, the drops in reaction time

Between positions one and fifteen for the permission and extension warnings were also negative, and statistically significantly so; ($\beta = -80.27$, $SE = 16.93$, $t = -4.74$, $p = <0.001$) and ($\beta = -50.15$, $SE = 18.13$, $t = -2.77$, $p = 0.006$) respectively. Because we have ruled out fatigue, we can infer that the negative slopes of the permission and extension warnings are indicative of generalization carrying over from the performance warning. However, these two warnings' slopes did not differ from one another $\beta = 421.5763$, $SE = 491.7897$ $df = 532$, $t = 0.857$, $p = 0.3917$, indicating that the rate of generalization was, while constant, nondiscriminatory. In contrast, the slopes for the save-executable and open-macro warnings were not different from zero; ($\beta = -37.81$, $SE = 16.38$, $t = -2.31$, $p = 0.021$) and ($\beta = 4.29$, $SE = 16.38$, $t = 0.26$, $p = 0.793$) respectively. This is consistent with the mouse cursor tracking results. Because the last two warnings which were were quite visually discrepant from the performance notification did not have statistically different reaction times between positions 15 and 1, and because the first two warnings which were quite visually similar to the

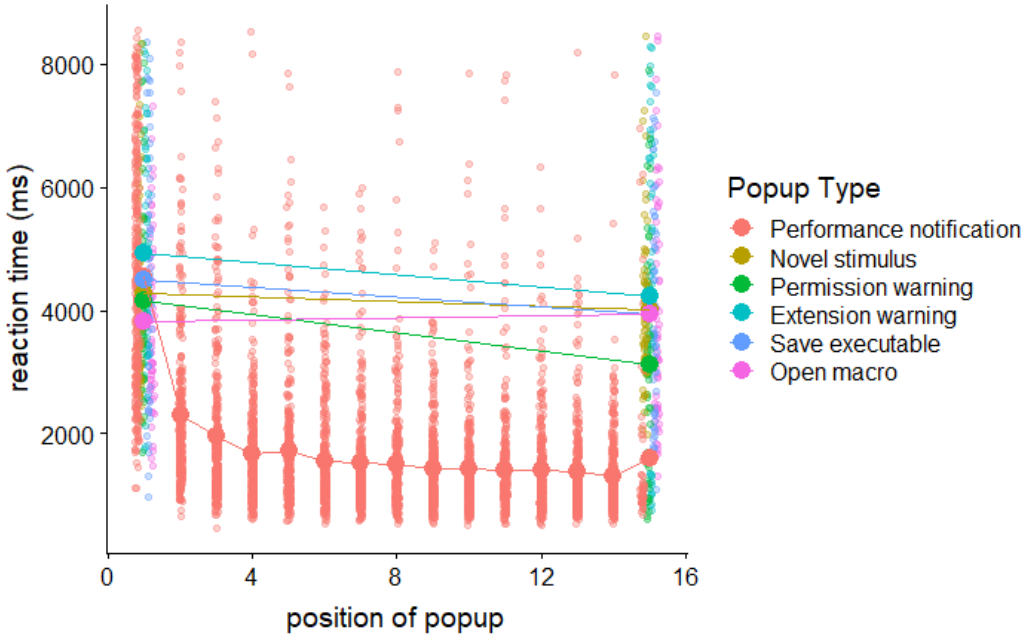



**Figure 12. Reaction times for each warning at various positional appearances.**

**Table 6. Predicting reaction time by interaction of modal position and modal type. 0-intercept for ease of interpreting the slopes. Practical effects of slopes (ms reaction speeds at position 15) are obtained by multiplying the estimate by 15 and adding to the corresponding main effect.**

| | reaction time | | | |
|---|---|---|---|---|
| ***Predictors*** | *Estimates* | *std. Error* | *Statistic* | *p* |
| **Performance** | 2720.52 | 36.87 | 73.79 | <0.001 |
| **Novel stimulus** | 4240.49 | 181.31 | 23.39 | <0.001 |
| **Permission warning** | 4257.72 | 184.81 | 23.04 | <0.001 |
| **Extension warning** | 4970.84 | 196.76 | 25.26 | <0.001 |
| **Save executable** | 4536.18 | 174.85 | 25.94 | <0.001 |
| **Open macro** | 3850.46 | 174.85 | 22.02 | <0.001 |
| **Performance × warning position** | -124.13 | 4.26 | -29.14 | <0.001 |
| **novel stimulus × position** | -13.20 | 16.76 | -0.79 | 0.431 |
| **Permission warning × position** | -80.27 | 16.93 | -4.74 | <0.001 |
| **Extension × position** | -50.15 | 18.13 | -2.77 | 0.006 |
| **Executable × position** | -37.81 | 16.38 | -2.31 | 0.021 |
| **Open macro × position** | 4.29 | 16.38 | 0.26 | 0.793 |
| **Observations** | 5586 | | | |
| **$R^2$ / adjusted $R^2$** | 0.757 / 0.756 | | | |

performance warning had statistically faster response times at position 15 than 1, these findings support the hypothesis that similar look-and-feel of security warnings to other notifications may be trigger generalization.

**Survey responses**

In a post-task survey (included in the appendix), participants reported the concern they felt when they encountered their assigned security warning. On the whole, participants reported anticipated levels of concern for the messages. Higher levels of concern were reported for security warnings, including the open-macro warning, permission warning, and save-executable warnings, whereas low concern was reported when seeing the novel stimulus or the performance notification. This pattern held for participants who saw the messages at either the first or the fifteenth position (see Figure 13).

We also asked participants for their preferred operating system, preferred web browser, whether they noticed seeing their assigned security message (a manipulation check), their general risk perceptions, and their information security threat severity and susceptibility perceptions. By and large, our participants preferred Windows (82.4%, *n*=551) over Mac (14.6%, *n*=98) or "other" (0.03%, *n*=20). Participants were neatly split between preferring Firefox and Chrome (48.7%, n=326 and 46.8%, n=313 respectively), with a sprinkling of other participants preferring Edge (n=7), Safari (n=13), Opera (n=6), or "other" (n=4). Participants in general reported above-average risk-taking attitudes (mean=5.61, SD=1.41), above-average perceptions of severity of a personal information security attack (mean=5.38, SD=1.47), yet lower perceptions of susceptibility to information security attacks (mean=4.16, SD=1.46) (each reported mean is an aggregate of three 7-

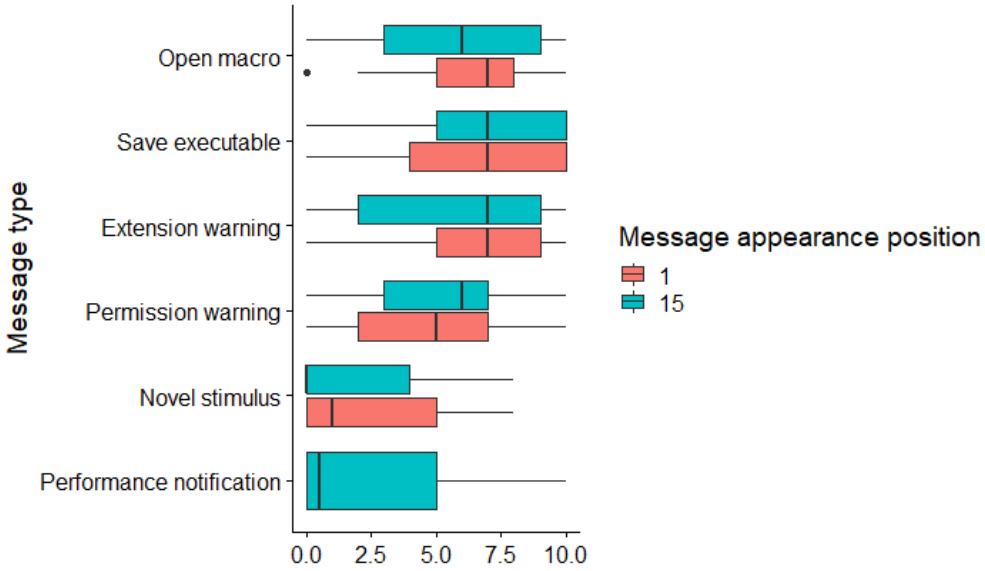



**Figure 13. Concern for message (self-reported, scale of 0 to 10).**

point Likert-scale agree-disagree survey items for each construct).

ANOVAs were performed for each survey construct individually to test whether responses were predictive of whether a security warning was clicked-through. The only significant overall ANOVA F-statistic was for the manipulation check ($F=28.997$, $p < .001$). A follow-up pairwise analysis with Tukey adjustment for each security warning grouped by appearance position suggested that participants who saw the extension install security warning at position 15 were more 16.28 times more likely (SD = 2.89) to have not noticed it than were participants who saw either the Open Macro security warning or the Save File security warning message at position 15 Also at position 15, participants approached being statistically more likely to have failed the manipulation check for the Extension security warning than for the Location Permission one (two-tailed $p = 0.064$). No pairwise comparison at position 1 was statistically significant. These findings provide some support for the notion that participants were less likely to notice (were more likely to have generalized habituation) to security warnings more visually similar to the performance notification after 14 exposures to the latter, than to less visually similar ones.

## 5. Discussion

This study contributes by showing the conditions under which generalization of habituation from routine notifications to security warnings occurs. Our paper does not claim to be the first to report the confusing of one warning with another [3, 15]. In contrast, our study specifically measures and tests the occurrence of generalization, and shows under what conditions it occurs.

Similarly, although our previous work has studied habituation in depth, we have not examined how habituation to one warning generalizes to another. Further, we know of no study besides the present study that investigates how habituation to a non-security-related notification can generalize to security warnings.

This paper (1) specifically examine how visual similarity leads to generalization, (2) test how habituation to a notification can generalize to different types of warnings, and (3) rule out the rival explanation of fatigue.

Specifically, we contribute by showing the following:

1. We provide empirical evidence that habituation to a frequent non-security-related notification does generalize to a one-time security warning.
2. We measure generalization in terms of (a) decreased attention to warnings, both in mouse cursor speed and response time; and (b) lower warning adherence behavior.
3. We show that this carry-over effect is due to generalization, and not fatigue. In past habituation literature, habituation and fatigue have been considered to be more or less synonymous (e.g., [1, 2]), but they are distinct phenomena with different implications. We show that participants ignored warnings not because they were tired, but because they had previously habituated to the performance notifications.
4. Finally, our results demonstrate that not all security warnings are equal in terms of the amount of generalization of habituation. Our results indicate that the more similar the security warning is to the non-security warnings in terms of "look and feel", the greater the degree of generalization. This finding questions whether corporate efforts to create a consistent UI look and feel is promoting better security or inhibiting security.

These insights open new avenues of research, pointing the way for researchers and practitioners to develop and test security warning designs that are resistant to generalization by distinguishing the appearance of security warnings from common notifications.

## 6. Limitations and Future Research

Our research was subject to several limitations. First, this research examines how similarity of appearance between notifications and security warnings can lead to the occurrence of generalization. Future research can additionally examine whether changing the mode of interaction for security warnings from the common "click to dismiss" paradigm can also reduce generalization.

Second, our experiment was designed to expose participants to notifications at a higher rate than is normally encountered in the same amount of time during usual computer usage. In future research, it would be interesting to explore if generalization of habituation occurs with the same amount of exposures distributed across a longer time window. However, participants' exposure to up to 15 notifications during the experimental session is not that far off from the number of notifications reported in observational studies [26]. Similarly, although the warning messages were meant to appear as if they were triggered by the website for each image, some messages (e.g., the save executable message) may have appeared incongruent for the experimental task. Consequently, some users may have been more dismissive than if the warning message better matched the task context.

Finally, while we explicitly controlled for fatigue in our experimental design, there are other factors that could have affected the speed and accuracy of participants' responses in our task. For example, participants could have become more engrossed in the task over time and therefore been quicker to dismiss notifications and less accurate at responding to warnings. Alternatively, faster responding may have been

due to participants learning about the task (e.g., which locations to click and when). For this reason, future work will be needed to tease out these alternative explanations. While habituation is a type of learning, it involves different low-level neural mechanisms than higher-order skill learning processes. Because habituation is fundamentally a neurobiological phenomenon, neurophysiological tools such as EEG or fMRI, may be especially useful to tease out these alternative explanations.

## 7. Conclusion

Generalization of habituation is a serious problem because it may cause users to tune out important security notifications, even if it is the first time any particular notification is displayed. However, an awareness of this problem can encourage software developers to create visually novel notifications that will receive the requisite attention to facilitate users' adherence to security warnings.

## ACKNOWLEDGMENTS

This was supported by the National Science Foundation under Grant CNS-1931108.

## REFERENCES

[1] M.E. Acer, E. Stark, A.P. Felt, S. Fahl, R. Bhargava, B. Dev, M. Braithwaite, R. Sleevi and P. Tabriz. 2017. Where the Wild Warnings Are: Root Causes of Chrome HTTPS Certificate Errors. in *ACM Conference on Computer and Communications Security (CCS)*, Dallas, TX.

[2] D. Akhawe and A.P. Felt. 2013. Alice in Warningland: A Large-Scale Field Study of Browser Security Warning Effectiveness *Proceedings of the 22nd USENIX Conference on Security*, USENIX Association, Washington, D.C., 257-272.

[3] T.S. Amer and J.-M.B. Maris. (2007). Signal words and signal icons in application control and information technology exception messages—Hazard matching and habituation effects. *Journal of Information Systems*, 21 (2). 1-25.

[4] B.B. Anderson, J. Jenkins, A. Vance, C.B. Kirwan and D. Eargle. (2016). Your Memory is Working Against You: How Eye Tracking and Memory Explain Susceptibility to Phishing. *Decision Support Systems*, 92. 3-13.

[5] B.B. Anderson, C.B. Kirwan, J.L. Jenkins, D. Eargle, S. Howard and A. Vance. 2015. How Polymorphic Warnings Reduce Habituation in the Brain: Insights from an fMRI Study *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ACM, Seoul, Republic of Korea, 2883-2892.

[6] B.B. Anderson, C.B. Kirwan, J.L. Jenkins, D. Eargle, S. Howard and A. Vance. 2015. How Polymorphic Warnings Reduce Habituation in the Brain: Insights from an fMRI Study. in *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, Seoul, South Korea, ACM.

[7] B.B. Anderson, A. Vance, C.B. Kirwan, J. Jenkins and D. Eargle. (2016). From Warnings to Wallpaper: Why the Brain Habituates to Security Warnings and What Can Be Done about It. *Journal of Management Information Systems*, 33 (3). 713-743.

[8] Apple.com. 2017. https://developer.apple.com/design/.

[9] R. Böhme and S. Köpsell. 2010. Trained to Accept?: A Field Experiment on Consent Dialogs. in *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)* Atlanta, ACM, 2403-2406. 10.1145/1753326.1753689

[10] C. Bravo-Lillo, L. Cranor, S. Komanduri, S. Schechter and M. Sleeper. 2014. Harder to Ignore? Revisiting Pop-Up Fatigue and Approaches to Prevent It. in *10th Symposium On Usable Privacy and Security (SOUPS 2014)*, USENIX Association, 105-111.

[11] C. Bravo-Lillo, S. Komanduri, L.F. Cranor, R.W. Reeder, M. Sleeper, J. Downs and S. Schechter. 2013. Your Attention Please: Designing Security-decision UIs to Make Genuine Risks Harder to Ignore *Proceedings of the Ninth Symposium on Usable Privacy and Security* ACM, Newcastle, United Kingdom, 1-12.

[12] J.C. Brustoloni and R. Villamarín-Salomón. 2007. Improving Security Decisions with Polymorphic and Audited Dialogs. in *Proceedings of the Third Symposium on Usable Privacy and Security (SOUPS 2007)*, New York, NY, USA, ACM, 76-85.

[13] A. Byers and J.T. Serences. (2012). Exploring the relationship between perceptual learning and top-down attentional contro. *Vision Research*, 74. 30-39.

[14] A. Cooper, R. Reinmann and D. Cronin. 2007. *About Face 3: The Essentials of Interaction Design*. Wiley, Indianapolis, IN.

[15] S. Egelman, L.F. Cranor and J. Hong. 2008. You've Been Warned: An Empirical Study of the Effectiveness of Web Browser Phishing Warnings *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, Florence, Italy, 1065-1074.

[16] R.L. Goldstone. (1998). Perceptual Learning. *Annual Review of Psychology*, 49 (1). 585-612.

[17] C. Green and D. Bavelier. (2003). Action video game modifies visual selective attention. *Nature*, 423. 534-537.

[18] M. Harbach, M. Hettig, S. Weber and M. Smith. 2014. Using personal examples to improve risk communication for security & privacy decisions *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, ACM, Toronto, Ontario, Canada, 2647-2656.

[19] M.J. Kalsher and K.J. Williams. 2006. Behavioral Compliance: Theory, Methodology, and Results in Wogalter, M. ed. *The Handbook of Warnings*, CRC Press, 313.

[20] C.B. Kirwan and C.E.L. Stark. (2007). Overcoming interference: An fMRI investigation of pattern separation in the medial temporal lobe. *Learning & Memory*, 14 (9). 625-633.
[21] S. Krug. 2014. *Don't Make Me Think, Revisited: A Common Sense Approach to Web and Mobile Usability*. New Riders.
[22] Microsoft. 2017. https://www.microsoft.com/en-us/design.
[23] C.H. Rankin, T. Abrams, R.J. Barry, S. Bhatnagar, D.F. Clayton, J. Colombo, G. Coppola, M.A. Geyer, D.L. Glanzman, S. Marsland, F.K. McSweeney, D.A. Wilson, C.-F. Wu and R.F. Thompson. (2009). Habituation revisited: An updated and revised description of the behavioral characteristics of habituation. *Neurobiology of Learning and Memory*, 92 (2). 135-138. http://dx.doi.org/10.1016/j.nlm.2008.09.012
[24] R.W. Reeder, A.P. Felt, S. Consolvo, N. Malkin, C. Thompson and S. Egelman. 2018. An Experience Sampling Study of User Reactions to Browser Warnings in the Field *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Montreal QC, Canada, 1-13.
[25] F. Schaub, R. Balebako, A.L. Durity and L.F. Cranor. (2015). A Design Space for Effective Privacy Notices. *To appear in the*.
[26] A.S. Shirazi, N. Henze, T. Dingler, M. Pielot, D. Weber and A. Schmidt. 2014. Large-scale assessment of mobile notifications. . in *SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*, Toronto, Ontario, Canada, ACM, 3055-3064. 10.1145/2556288.2557189
[27] A. Sotirakopoulos, K. Hawkey and K. Beznosov. 2011. On the Challenges in Usable Security Lab Studies: Lessons Learned From Replicating a Study on SSL Warnings *Proceedings of the Seventh Symposium on Usable Privacy and Security (SOUPS)*, ACM, Menlo Park, CA, 3:1-3:18.
[28] Z.R. Steelman, B.I. Hammer and M. Limayem. (2014). Data collection in the digital age: innovative alternatives to student samples. *MIS Quarterly*, 38 (2). 355-378.
[29] J. Sunshine, S. Egelman, H. Almuhimedi, N. Atri and L.F. Cranor. 2009. Crying Wolf: An Empirical Study of SSL Warning Effectiveness. in *SSYM'09 Proceedings of the 18th Conference on USENIX Security Symposium*, Montreal, Canada, 399-416.
[30] Symantec. 2017. Internet Security Threat Report.
[31] R.F. Thompson and W.A. Spencer. (1966). Habituation: A Model Phenomenon for the Study of Neuronal Substrates of Behavior. *Psychological Review*, 73 (1). 16-43.
[32] A. Vance, B. Brinton Anderson, C. Brock Kirwan and D. Eargle. (2014). Using Measures of Risk Perception to Predict Information Security Behavior: Insights from Electroencephalography (EEG). *Journal of the Association for Information Systems*, 15 (10). 679-722.
[33] A. Vance, J. Jenkins, B. Anderson, D. Bjornn and B. Kirwan. (2018). Tuning Out Security Warnings: A Longitudinal Examination of Habituation through fMRI, Eye Tracking, and Field Experiments. *MIS Quarterly*, 42 (2). 355-380.
[34] A. Vance, C.B. Kirwan, D. Bjornn, J.L. Jenkins and B.B. Anderson. 2017. What Do We Really Know about How Habituation to Warnings Occurs Over Time? A Longitudinal fMRI Study of Habituation and Polymorphic Warnings *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, Denver, CO.
[35] R. Wash, E. Rader and C. Fennell. 2017. Can People Self-Report Security Accurately?: Agreement Between Self-Report and Behavioral Measures *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, Denver, CO, USA, http://dx.doi.org/10.1145/3025453.3025911.
[36] J. Weinberger and A.P. Felt. 2016. A week to remember: The impact of browser warning storage policies. in *Symposium on Usable Privacy and Security (SOUPS)*, Denver.
[37] R. West. (2008). The Psychology of Security. *Communications of the ACM*, 51 (4). 34-40.
[38] M.S. Wogalter. 2006. Communication-Human Information Processing (C-HIP) Model. in Wogalter, M.S. ed. *Handbook of Warnings*, Lawrence Erlbaum Associates, Mahwah, NJ, 51-61.

## Appendix A – Post-task Survey

Please select your gender:

- Female
- Male
- Other

Please enter your age: ____

Please select your preferred OS:

- Mac
- Windows
- Other

Please select your preferred browser:

- Chrome
- Edge
- Firefox
- Opera
- Safari
- Other

*Presentation order for the following items was randomized. All items in this section allowed respondents to choose from the following Likert-scale options:*

- *1-Strongly disagree (1)*
- *2-Moderately disagree (2)*
- *3-Mildly disagree (3)*
- *4-Neutral (4)*
- *5-Mildly agree (5)*
- *6-Moderately agree (6)*
- *7-Strongly agree (7)*

[RISK1] Ignoring malware warning screens can cause damages to computer security.

[TSUS1] My computer is at risk for becoming infected with malware.

[RISK2] Ignoring malware warning screens can put important data at risk.

[TSUS2] It is likely that my computer will become infected with malware.

[TSUS3] It is possible that my computer will become infected with malware.

[RISK3] Ignoring malware warning screens will most likely cause security breaches.

[TSEV1] If my computer were infected by malware, it would be severe.

[TSEV2] If my computer were infected by malware, it would be serious.

[TSEV3] If my computer were infected by malware, it would be significant.

[attention check] Select "3-mildly disagree" for this answer (attention).

*The following questions appeared at the end of the survey:*

[manipulation_check] Did you notice the following popup during the Batman image classification task?

[Yes / No/ I'm not sure]

[realism] On a scale of 0 to 10, how realistic do you think the following message is? [participants were shown a screenshot of the security notification for their treatment group]

[0-Not realistic (1) ... 10-100% realistic (11)]

[concern] On a scale of 0 to 10, how concerned did the following screen make you feel during the Batman image classification task? [participants were shown a screenshot of the security notification for their treatment group]

[0-Not concerned at all (1) ... 10-Extremely concerned (11)]

[debrief] The primary objective of this study was to observe how you responded to browser popups. You were randomly assigned to a condition in which you saw a variant of a browser popup. Additionally, the browser popups you saw were simulated. Your response to them will have no impact on your browser or computer.

[free_response] Any feedback for the research team?

[free response]

# "There is nothing that I need to keep secret": Sharing Practices and Concerns of Wearable Fitness Data

Abdulmajeed Alqhatani
*College of Computing and Informatics*
*University of North Carolina at Charlotte*
*aalqhata@uncc.edu*

Heather Richter Lipford
*College of Computing and Informatics*
*University of North Carolina at Charlotte*
*Heather.Lipford@uncc.edu*

## Abstract

There has been increasing use of commercial wearable devices for tracking fitness-related activities in the past few years. These devices sense and collect a variety of personal health and fitness data, which can be shared by users with different audiences. Yet, little is known about users' practices for sharing information collected by these devices, and the concerns they have when disclosing this information across a variety of platforms. In this study, we conducted 30 semi-structured interviews with wearable fitness device users to understand their sharing intentions and practices, and to examine what they do to manage their privacy. We describe a set of common goals for sharing health and fitness information, which then influence users' choices of the recipients and the specific practices they employ to share that information. Our findings indicate that participants were primarily concerned about acceptable norms and self-presentation rather than the sensitivity of the information. Our results provide a set of common goals and practices which can inspire new applications and help improve existing platforms for sharing sensed fitness information.

## 1. Introduction

Wearable sensing devices for health and fitness tracking have become ubiquitous. Researchers anticipate that fitness devices, including smart watches, will continue to lead the wearables market in future years [18]. Such devices provide users a variety of personal sensed data, such as step count, exercise, vital signs, and sleep quality. By using these metrics, users can become aware of their activity, thus improving their healthy practices. For instance, individuals who are overweight can use accelerometers to increase their daily steps and burn more calories. People can also utilize the features within fitness trackers to monitor some medical conditions (e.g. diabetes).



*USENIX Symposium on Usable Privacy and Security (SOUPS) 2019.*
August 11 -- 13, 2019, Santa Clara, CA, USA.

Sharing health and fitness information has also become an important part of many users' practices towards achieving their health and fitness goals. Thus, most wearable devices today have also social features that allow users to share their information and interact with different people and organizations. Some devices like Fitbit have built-in social circles where users can talk about their exercises, goals, and progress. Alternatively, users can broadcast their wearable fitness data on external health and fitness apps (e.g. Strava & RunKeeper), via common communication applications (e.g. WhatsApp), or over popular social media applications (e.g. Facebook & Twitter). In addition, users may share data with insurers or through workplace campaigns to receive rewards to further incentive healthy behaviors. Individuals may utilize several different platforms, and more than one communication channel, and switch between them to share their information online [30].

Researchers have primarily examined fitness device data sharing on social media platforms [17, 21, 24, 25] and in the workplace [3, 7, 8], revealing a range of common reasons and outcomes for sharing. Others have investigated privacy implications and concerns, including the sensitivity of various information and the lack of understanding of fitness trackers' data practices [22, 27]. We expand upon this work by investigating users' practices across the range of sharing that they perform. Thus, we examine both aspects of social privacy – how users disclose and interact with other people around their shared data, as well as data privacy issues that arise when they provide their data to additional organizations. As Contextual Integrity theory posits, information sharing is governed by the norms and expectations of the context, and privacy problems occur when those expectations are violated [19]. Thus, we examine those expectations to provide insight into the privacy concerns and needs of users of wearable fitness trackers.

More specifically, contextual integrity predicts that sharing preferences of people may vary based on the receiver, the information type, and transmission principle. Thus, our original goals were to examine how the design of the devices and sharing platforms, the sensitivity of certain kinds of data, and the availability of controls influence users' disclosure behaviors and privacy concerns. However, we found that participants' behaviors had less to do with such factors than with their sharing goals and the associated audiences

related to those goals. Thus, our study contributes a set of common sharing patterns, relating goals, audiences, and specific practices of participants as part of understanding users' privacy decision making and behaviors.

Specifically, we report on a qualitative study with 30 existing users of wearable fitness trackers, who have shared their information with different audiences, in order to understand their sharing practices and behaviors. Our results indicate that users' concerns about disclosing wearable fitness information are about self-presentation goals and acceptable behaviors of sharing with people on different platforms, rather than concerns over the sensitivity of data.

The contributions of our work are:

- We enhance the understanding of the sharing of wearable fitness data across the range of platforms and audiences of users.
- We provide a set of common patterns of sharing goals, audiences, and practices.
- We provide insights about users' perceptions regarding their disclosure decisions and privacy perceptions, and present implications for researchers and designers to help users share their information as desired.

## 2. Related Work

A number of studies have examined technologies for tracking and sharing fitness data, prior to the widespread adoption of wearable sensing devices. Such studies found that users appreciated the social aspects of sharing data within online communities and social media, and doing so encouraged them to pursue a healthy lifestyle [20, 25]. Such social interaction can provide accountability, advice, and emotional support for individuals' health and fitness goals [25]. However, users sometimes had trouble determining what to share, and with whom [15, 17]. In addition, users struggled to find desired sharing features on various platforms [15].

The introduction of wearable fitness trackers has contributed to widespread tracking and sharing of fitness information. Today, many modern wearable fitness devices have socially embedded features that allow users to share data with other people. One study found that half of their participants utilized social features of their devices to support their fitness activities [6]. Thus, a number of researchers have examined users' motivations and behaviors of sharing fitness tracker data on different platforms and within different domains [8, 9,11, 21, 24, 31], as well as users' privacy concerns and the privacy implications of sharing this information [4, 7], which we discuss in more detail below.

### 2.1 Wearable Fitness Data Sharing

Multiple studies have explored the opportunities of sharing wearable fitness data with people online, primarily on social media sites [9, 11, 21, 24, 31]. Studies reveal that users' motivations for sharing include increasing motivation through accountability, finding social support, and competing with others, all in support of a user's health or fitness goals [5, 11]. Indeed, studies have linked social sharing and competing with an increased intention to exercise [31]. Some users were also motivated to share in order to help and connect with others with similar goals; however, finding the right community that satisfies sharing needs can be challenging [6].

For example, Gui et al. performed a qualitative study of a fitness plugin for the Chinese social networking service WeChat [9]. Participants preferred utilizing their existing network of social contacts, and sharing fitness data as part of their regular social networking practices. Users found increased opportunities for interaction with others who also shared their fitness information. Participants also reflected on the impressions that can be inferred by others as a result of such sharing. However, sharing using such social features may lack emotional support, or can be less effective when sharing with unknown contacts [9].

A similar study by Dong et al. focused on the Chinese site Weibo [5], and found that users shared data from the wearable sensing device Mi Band primarily to record their life and motivate oneself. In comparing users' motivations for posting on Weibo versus WeChat, their findings also provide evidence that people have different motivations for sharing on different social network sites, based on the different audiences found on different sites [9]. As a result, people may integrate different social platforms into their communication practices in order to reach a broader audience to meet their intended goals [30].

Finally, several studies have investigated sharing data from activity-sensing devices within workplace settings in order to promote physical activity [3, 7, 8]. Employees often perceive workplace tracking and sharing programs to be beneficial to their goals [3]. Financial incentives were also seen as beneficial, and could be a motivator for participation [3]. Employees generally do not feel that step count is sensitive data and are willing to share with their workplace and colleagues; however, perceptions can change over time as people gain an understanding of how details of their personal lives and activities are reflected in just a step count [7]. Employees may feel pressure to explain low step counts or change their activities to fulfill team goals, and require additional negotiation of personal boundaries with colleagues [7]. People are also concerned that undesirable decisions can be made based on the wearable fitness data they share with their employers [4].

Thus, research has so far examined the sharing of wearable fitness data primarily on social media, or within the workplace, examining users' goals and perceptions. We expand upon this work by broadening our focus to all forms of sharing, to investigate more general patterns and factors that influence users' sharing decisions.

### 2.2 Privacy Concerns

Collecting and sharing health related data raises several privacy risks [29]. Researchers have examined the contextual nature of sensed data, both from wearables and a variety of IoT devices. Users' comfort in sharing information obviously depends greatly on the type of information, the recipient, and the reason and benefits for sharing [16, 23]. Several studies have found that users of wearable sensing devices do not find movement data such as step count as sensitive [10, 14], and thus people are willing to share such information with many different audiences. Weight and sleep are considered more sensitive, depending on the audience [12, 23]. The primary privacy concern is with locational information captured by devices with GPS. Users fear that location data shared on social media applications can be used by strangers, or even criminals, to know where they live [10, 14].

In a survey study, Lowens et al. [13] revealed two other privacy concerns about sensed fitness data: unintended use and lack of control of personal information. For instance, users are concerned that a health insurance company could get access to their fitness data and adjust their coverage's rate accordingly [13]. Users also expressed a strong desire to retain control and ownership over their data.

In many countries, regulations protect the collection, storage, and sharing of health-related data. Yet, it is unclear how the data from wearable tracking devices is covered by those regulations [1]. For example, Paul and Irvine [22] analyzed the privacy policies of four popular wearable fitness companies. They found that some of these services did not comply with existing privacy regulations regarding informing users about the use of health- related data. In recent work, Vitak et al. found that users of fitness trackers have limited knowledge of the data practices of fitness devices' manufacturers [27].

Thus, while research has highlighted several concerns of users over sharing certain information, we seek to understand more about the privacy-related decisions and practices of long-term users of wearable fitness trackers, and how those decisions are related to users' motivations for disclosing personal sensed data.

## 3. Methodology

### 3.1 Interview Study

We conducted 30 semi-structured interviews with wearable fitness users (15 males and 15 females) to examine their sharing and privacy preferences. As our focus is on the sharing of fitness information, we restricted participation to people who have a wearable fitness device for at least three months and who have shared their information recorded by the device with other people or organizations. Participants were first prompted to fill out an online screening survey. We then contacted the participants who met our criteria to schedule an interview. All the interviews except one were conducted remotely on the phone, over Skype, or Google Hangout. Interviews lasted on average 25 minutes, and ranged from 14 to 43 minutes. Study participants were all compensated with a $10 Amazon gift card for their time after completing each interview.

The interview questions were structured into three parts. The first part discussed the general usage of the wearable fitness devices by participants. We asked the participants how frequently they use their devices, and how and when they check their sensed data. In the second part, we focused on the users' practices and behaviors with respect to the sharing of their data. For example, participants were asked how they share their information, what information they share and with whom, and what platforms they utilize. Some participants were uncertain about their profiles, so we requested they go through their accounts to answer the previous questions. This was followed by questions about the participants' sharing practices, and the impact of sharing on their behaviors and use of their devices. In the last part, we asked the interviewees about their privacy concerns and how they manage their privacy. Participants mentioned different scenarios regarding how their information could be misused, and expressed several sharing and privacy needs. The full interview is listed in the Appendix.

### 3.2 Participants

Interview participants were recruited between April and June 2018. We recruited participants by posting flyers at fitness centers near our university campus, and by advertising on relevant Reddit forums. Our methods of recruitment allow us to have participants from diverse age groups and professions. The average age of our participants was 32 years old, ranging from 20 to 51. Educational backgrounds of the interviewees ranged from high school to doctorate. Table 1 reports the demographics of our participants along with the devices they have.

| | Gender | Age | Occupation | Device(s) |
|---|---|---|---|---|
| P1 | M | 38 | Software developer | Nokia |
| P2 | M | 29 | Physician | Apple Watch |
| P3 | M | 34 | Software Engineer | Garmin |
| P4 | M | 39 | Designer | Apple Watch |
| P5 | M | 44 | Self-Employed | Apple Watch |
| P6 | M | 47 | Risk Manager | Apple Watch |
| P7 | F | 29 | Food Services | Apple Watch |
| P8 | F | 28 | Designer | OMbra |
| P9 | M | 38 | Fire Fighter | Garmin |
| P10 | M | 51 | Computer Engineer | Garmin; Fitbit |
| P11 | F | 25 | Gerontology Researcher | Apple Watch |
| P12 | F | 27 | Event Rentals | Jawbone |
| P13 | M | 47 | Finance | Apple Watch |
| P14 | F | 31 | Marketing | Jawbone |
| P15 | M | 32 | Product Manager | Fitbit |
| P16 | F | 35 | Student | Fitbit |
| P17 | F | 35 | GIS Manager | Fitbit |
| P18 | M | 35 | Self-Employed | Apple Watch |
| P19 | F | 22 | Student | Fitbit |
| P20 | F | 26 | HR Manager | Fitbit |
| P21 | F | 27 | Student | Fitbit |
| P22 | F | 26 | Student | Fitbit |
| P23 | F | 32 | Teacher | Fitbit |
| P24 | F | 25 | IT | Apple Watch |
| P25 | M | 27 | Sales | Apple Watch |
| P26 | F | 20 | Student | Polar M600; Polar H10 |
| P27 | M | 27 | Software Administrator | Garmin |
| P28 | M | 48 | Journalist | Jawbone |
| P29 | M | 25 | Student | Nokia |
| P30 | F | 25 | IT project Manager | Motiv Ring |

**Table 1: Summary of the Participants' Information.**

### 3.3 Data Analysis

All interviews were audio-recorded and transcribed. We utilized an open coding approach using a qualitative data analysis tool to identify patterns from the participants' responses. Initially, three researchers analyzed three transcripts to develop a codebook, with discussions occurring between the researchers during this process. Coding saturation was met after coding these three transcripts, after which no more codes were added. The developed codebook consists of 26 codes. Each code was conceptually assigned to one of three categories: usage, sharing, or privacy. Then, the primary investigator and another researcher independently coded the remaining 27 transcripts using that codebook. The two coders kept track of their disagreements and the calculated inter-rater agreement was 80%. The remaining disagreements were discussed and resolved by the two coders. We note that any numbers reported in the results are not meant as quantitative analysis, but merely to indicate prevalence of themes in our sample of participants.

### 3.4 IRB Approval

To ensure the protection of human subjects, prior to the start of this study, our university Institutional Review Board (IRB) approved this study as an exempt protocol.

### 3.5 Limitations

Our study has limitations similar to many qualitative interview studies: a convenience sample of limited size that may not be generalizable to the broader population of users. The inclusion criteria for participation in our study required at least three months of device usage, which may not be enough to assess the sharing behaviors of the participants. Also, while we attempted to recruit users from diverse ages and professions and balanced participants with respect to gender, we did not consider their cultural backgrounds which may influence participants' views on sharing and privacy. Finally, in focusing on the broad range of participants' self-reported behaviors, interviewees may have neglected to report detailed or accurate sharing behaviors.

## 4. Results

Our participants utilize a variety of wearable health and fitness devices that have different sensors to track movements and vital signs (Table 1). The devices used in the study come in different form factors that include smart watch, chest strap, smart bra, and smart ring. Apple Watch is the most common device used by participants, followed by Fitbit. These two are the top-selling brands in the last two years [2]. A few participants have shared information from more than one wearable fitness device, but we excluded devices that have been used for less than three months. It is also noteworthy that Jawbone had gone out of business a few days before we completed our interviews. We begin with a general discussion related to participants' use and perspectives regarding their devices, before moving into more detail about sharing and privacy aspects.

### 4.1 Use: Motivations & Contexts

Participants reported several goals for using wearable health and fitness devices. Tracking physical activity, mainly step count, as well as being aware of general health were the primary reasons for use by all participants. Many of our interviewees have sedentary jobs, and they used the devices as a reminder to move. In addition, people make use of wearable fitness trackers to motivate themselves to exercise and to stay accountable. Aside from fitness tracking, a considerable number of participants reported using the devices for medical reasons, such as for recovery after surgery:

*"I had back surgery in October, so I use it as a tool to make sure that I am maintaining my recovery from my surgery"* (P11).

For many participants, the impact of using a device is measured by whether or not goals are attained or behaviors have changed. For example, four interviewees who wanted to lose body weight expressed positive attitudes toward the device because it supports them by tracking how many calories they consume and how many pounds they lose every week. Other participants used a wearable tracker to monitor vital signs, such as heart rate, or to track sleep quality. For instance, P22 has sleep apnea and she uses a Fitbit to assist her in detecting the problem. Unlike human beings, a wearable device provides unemotional facts about one's health status. P10 stated: *"The device is sort of truth because humanly you can say I have an active day, I have a busy day. In actual fact you were busy on your desk, whereas the fitness device is unemotional. It's unaware of how you're feeling. It's only aware of your physical movement."* In contrast, two participants did not find their wearable trackers to be helpful in achieving their fitness goals. For example, P6 believed that the device did not change his behavior, and he did not find features like badges and rewards within the device to be encouraging. Another participant indicated that the device was motivating when he first bought it but that impact has diminished over time, especially after his best friend who he used to exercise and share information with moved away. In general, the majority of participants were pleased about their wearable trackers and they stated several benefits that they received from their use.

We asked participants why they decided to use the device they have rather than a different device. As expected, Apple Watch was preferred due to its variety of metrics as well as its capability to integrate fitness tracking with other features, such as sending text messages and taking phone calls. Two participants who had Jawbone liked its design that encourages users to keep active by achieving scores. Other devices were chosen for other goals, such as heart rate monitoring. We also found that a single device may not fulfill some users' needs; as a result, they incorporate more than one device into their practice. For example, P26 reported using a smart watch to track her runs and to map routes during soccer games, as well as a chest strap to track heart rate. Similarly, P10 uses one device for running analytics and another one for general health data monitoring. While the wearable devices used by participants have different measurements, all have sensors to capture steps taken.

Participants also expressed different contexts for use. Most of the participants mentioned that they use their wearable tracker at all times even while sleeping. Five interviewees indicated that they use their devices at certain times, mostly when they are exercising or during an activity such as biking. Users reported reviewing regularly, either after a particular activity or in the morning or night to check regular nightly or daily statistics.

### 4.2 Patterns of Goals & Audiences

Participants' goals for sharing wearable fitness data are similar to those reported in other studies, such as competing with peers, mutual support, and boasting [5, 11]. However, we expand on previous research by describing a set of common sharing goals, audiences, and practices. Overall, we found that users make decisions about their audiences based on their goals, which also drive their choices regarding the way they communicate their wearable fitness data. Table 2 summarizes the common goals we found in our study.

Participants' goals were related to their choices of audience to help them with those goals. Our analysis revealed six categories of audiences: friends, family, strangers, physicians, financial incentive programs, and co-workers.

Our participants shared their information with friends (25/30), family (17/30), or both (13/30). Eleven participants indicated that they shared with strangers, mainly through different health and fitness forums. Sharing wearable fitness data with physicians for medical tracking was mentioned by seven participants; while five interviewees have their devices connected with third party applications such as insurance companies and pharmacists in order to receive financial discounts or rewards. Finally, four participants identified co-workers with whom they share data. Participants disclose more or less information depending on the recipients and their goals, with practices specific to those goals and audiences. Again, Table 2 summarizes these practices and we now discuss the details of those patterns.

#### 4.2.1 Friends

The majority of participants shared fitness data collected by wearable trackers with friends, often on social network sites. Accountability was mentioned as a strong motivation for sharing with friends. Participants also indicated that being able to see friends' activity progress and receive notifications about others' achievements encouraged them to pursue their fitness goals. Sharing can sometimes turn into competition and the desire to outperform each other by being the most active person in the day. Moreover, individuals may feel embarrassed if they failed to meet their fitness goals:

*"It's kind of partially just a motivational thing but also partially… I guess you can say it's kind of shame like that you know they see if you haven't set or hit your goals."* (P15).

Another goal that emerged from interviewees' responses for sharing wearable fitness data with friends was the intention

| Goals | Targeted Audience | Practices |
|---|---|---|
| Accountability<br><br>Competition<br><br>Boasting a positive self-image | Friends | • Share common sensed data only (e.g. step count).<br>• Sharing mostly done on social media channels.<br>• Share after good physical performance. |
| Support family maintaining a healthy lifestyle<br><br>Mutual & emotional encouragement | Family | • Disclose more information to family than to friends.<br>• Simple ways to communicate wearable data outside of platform |
| Feedback from experienced individuals<br><br>Accountability | Unknown people (strangers) | • Share using device built-in social communities, or on social media communities<br>• Share variety of non-identifiable information related to fitness goals |
| Vital signs monitoring<br><br>Tracking medical conditions (e.g. sleep apnea) | Physicians | • Disclose everything accurately<br>• Compile data manually, or show doctors data in the app. |
| Receive financial discounts/rewards | Insurance companies; Pharmacist; Employers | • Wear the device continuously to maximize the metrics<br>• Provide permission to incentive programs to access data directly on the device, or make sure to update data regularly. |
| Competition in the workplace | Co-workers | • Share step count only.<br>• Set regular step goals and interact with others to achieve. |

**Table 2: Participants' Goals & Practices Based on Audience.**

to boast and communicate a positive image about one's fitness and health. For a few people, accountability can only be met if other users acknowledge good physical performance.

However, sharing with peers for accountability may also impose challenges. A few participants, especially those who may not always have the time to exercise, expressed fears about friends' judgements of their lack of activity. Two participants also did not like to share fitness information with friends on social media because it might be perceived as bragging. Another participant decided to stop sharing with Facebook friends, and instead limited the sharing to a few friends with similar interests on the device's platform:

*"I kind of started feeling a lot of pressure when I was sharing it because I thought like, oh well if I share on Facebook and I don't share anything for a while, what everybody will think, or they gonna think that I stopped working out. And I think that for me this is on the perfect balance that I can share with my friends on the app and my friends on the app who are active can share with me"* (P22).

Others faced concerns over the broad audience on social media platforms, and limited the data they shared accordingly. For example, P12 sometimes found sharing wearable fitness data inappropriate; she stated: *"Facebook friends include co-workers or professional contacts, and it just seemed weird to share my fitness activity with people that I work with or people that I have a contact with them for professional reasons."* Several other participants chose to not share with friends on social media channels because of perceived lack of interest of their friends on those platforms.

All participants who shared with friends reported sharing basic sensed data, such as step count and distance covered. None of those participants disclosed more personal information with friends, such as body weight, and the majority of the interviewees were unwilling to share their eating habits. In addition, other detailed health data, such as heart rate or blood pressure, were not shared because such data was considered less interesting to friends.

Participants mostly utilized the features within their devices' apps to hook up their data in the trackers with their social media accounts. However, participants are selective on when and what to share in order to maintain a positive self-image. For example, instead of sharing on a regular or automated basis, they only shared data after positive physical performance.

#### 4.2.2 Family

Another audience that many participants mentioned sharing with is those they are closest to, primarily family members (e.g. spouse) and occasionally very close friends. In this case, sharing was more about mutual and emotional encouragement. Participants expressed feeling responsible to share their information or any experience they had, whether good or bad, to motivate their loved ones towards a healthy lifestyle:

*"We did have bad habits when it comes to food, and so I show them look at what happened when I was going*

*through depression on October. Look at how I was eating and look at my heart rate, and look at it right now. You know I share with them to show them, it is like you are family, you are just like I am"* (P5).

P13 commented that he does not usually workout with his wife, but he liked the challenges and notification features generated by the device, which makes it feel as if they are exercising together. However, only three participants utilized features within devices to share data with family members or a few close contacts. The majority reported using simple techniques to communicate their sensed data. For example, they simply talk about their activity goals and progress or show family members steps count in their trackers. We believe this is because family members may not have the same device to connect directly with the user.

Family members are typically aware of each other's health conditions; thus, it is not surprising that participants disclose more information to family than to friends. They reported feeling comfortable sharing personal information, such as weight or fitness goals, with family:

*"If I share it with more people, I would have chosen which specific pieces of information but I will still share everything with my wife"* (P13).

#### 4.2.3 Strangers

More than one third of the participants shared wearable fitness data with unknown people, mainly on fitness forums. Some used the communities on the device's platform, others found forums and groups on various social media sites, such as Reddit and Facebook. By sharing on these fitness forums, participants seek to receive help and feedback from experienced people (e.g. coaches) regarding specific fitness goals, such as weight loss. Holding oneself accountable was also a primary goal, through interacting with other people with similar interests and goals.

For example, P12 described herself as *"conservative"* with respect to sharing personal information. She used to share her fitness information with her friends on Facebook, but later felt uncomfortable because of a mismatch between her and her friends' interests. This participant then joined a women's fitness group on Facebook and restricted sharing to that group of strangers with similar interests. P22 also found a fitness group on Facebook and stated such sharing can be an opportunity to build relationships with others with similar goals. Another participant even reported that she is looking for a new device with better support for fitness communities, in order to connect with others with similar goals.

Unlike with friends, participants did report sharing body weight, calories consumed, and the type of food and exercise, in addition to step count. Participants expressed willingness to share because they saw little harm that could come to them:

*"I guess I share more even when I don't reach my goals because I want to… I don't know, because they are also on the same journey so I feel like it is for accountability and they are not going to use that information in a way that would negatively affect me."* (P12).

However, interviewees were unwilling to disclose data they saw as personal, such as location information, with unknown people due to safety concerns. A few also reported putting fake information in their accounts to protect their identities. In addition, they were less interested in sharing any sensed data that were considered irrelevant to their health or fitness goals.

#### 4.2.4 Physicians

We were surprised by how many participants also shared their data with doctors or other caregivers. Participants' intentions were to share vital signs with doctors, often due to medical conditions. For example, P22 had been working with her doctor to lose weight by sharing steps taken. In addition, she takes medicine that affects her heart rate functions and uses the device to monitor any heart rate abnormality. She also has sleep apnea and used the sleep logs feature to show her doctor her sleep quality. Another participant (P11) had back surgery and shared her data with physicians to keep track of her walking progress afterwards.

Participants indicated they were comfortable disclosing their data openly with doctors because it would be helpful to manage their health with accurate information. For instance, P18 stated: "*Generally, it made me more diligent in my recordings. I want to get things right if I am showing my doctor the information; it is accurate, and it is not misleading to my professionals.*"

Participants expressed frustration with the methods they utilize to communicate their wearable fitness data to physicians. All those participants, except one, reported that they manually record or copy data from the device's website into files, take a screenshot of data, or show their doctors the data in the app. They expressed desire for a centralized control where wearable device data can be integrated with other health information systems such as Electronic Health Records (EHRs) to allow medical providers to directly access their data and interact with them more easily.

*"I can't just share the data directly with my doctor. I have to compile the data and then present a report to my doctor, and that can be frustrating and time consuming."* (P18).

#### 4.2.5 Financial incentive programs

Our results reflect those reported in other studies (e.g. [3]) that users of wearable fitness devices may disclose their

sensed data in order to receive financial discounts or rewards. Interviewees found financial incentive programs to be a great motive to increase physical activity. Participants reported different recipients of their data for this goal, including insurance companies, employers, and pharmacists. Incentives can be received as prizes offered by an employer, or discounts on purchases and insurance rates. Participants update the data, mainly step count, on a regular basis through an employer's portal. Others provided permissions to incentives programs to pull the pedometer data automatically from their devices. In order to receive financial incentives, participants make sure to have their trackers on all the time to collect the data.

Two participants admitted that their primary goal for sharing was to receive financial incentives. P28 indicated that he connected his Jawbone to a pharmacy app in order to collect points based on the number of steps taken, which then can be redeemed as discounts on purchases. Similar to most of our participants, this user had no concern about sharing this type of data. He stated: *"The sharing with the pharmacy, there has been a very motivating financial affect, maybe 20 dollars every few months. It is free money, but I never been giving away something confidential. I was giving away the number of steps or my weight. There is nothing that I need to keep secret."* He repeatedly described the sharing experience on the device as a "game" where one tries to achieve high scores.

Another participant, P15, linked his Fitbit account to his employer health insurance portal. He considered himself healthy right now but was worried about the possibility of increasing his premium based on his fitness condition in the future.

Two other participants did not share their information for any financial incentives, but expressed a desire to share if they were offered this option:

*"If it is something that gives me a discount, I will definitely share information with them. I probably will be more inclined to. It gives me another reason to be active to save money on my health insurance."* (P25).

#### 4.2.6 Co-workers

Finally, a few people identified co-workers with whom they disclosed wearable fitness information as part of participating in workplace health campaigns. Some organizations offer employees the option to link their trackers' data to the employer's website. For many of those participants, sharing with co-workers is a "*friendly competition,"* although prizes can sometimes be offered to further motivate participants. P3 stated: *"When I originally started wearing it's because of the competition. You don't wanna be at the bottom of the list of your co-workers so you wanna be more active."* However, participants find sharing fitness data with workmates as an opportunity to increase physical activity, especially because some of these participants have jobs that restrict their physical movements during the day. It can also be an opportunity to reinforce behavior change, so moving and exercising become habits rather than merely competition. Participants set daily step goals and send cheers to other co-workers who hit their step goals. To compare data with other co-workers, participants sync their daily step count to the employer's system. Although some of the workmates may not personally be known, sharing the number of steps walked every day was not something they were concerned about.

### 4.3 Sharing Impact

We asked participants about their perspectives regarding sharing and how it impacts their behaviors. Most of the participants (19/30) said that sharing wearable fitness data has impacted them in a positive way. It helped participants to become more aware about their health and fitness status and encouraged them to stay competitive and accountable. Similar to [23], we found that users' sharing behaviors may change over time, especially with respect to the level of information shared. For example, one participant realized that she became willing to share more data in order to motivate herself to exercise:

*"I would like to share more because I noticed that the more that I'm sharing with people that I feel comfortable with I guess or with people that are having the same goal, the more I feel I exercise more"* (P12).

In contrast, another participant decided to share less information with the public, mainly because of privacy concerns:

*"I actually think I share publically a lot less than I used to because I become more concerned with privacy, but I have been sharing more information with some of my private connections"* (P18).

Five interviewees were uncertain about the impact of sharing. These participants indicated that sharing wearable fitness data has provided some benefits, such as the desire to exercise, but it did not help them to achieve desired goals. In addition, participants with mixed feelings indicated that the impact depends on the audience's reaction and feedback.

Another six participants stated that sharing did not impact their behaviors. Some of these interviewees commented that they are self-motivated, but they shared their wearable fitness data to help others and for enjoyment:

*"I just enjoy sharing the information and posting the challenge to my followers to keep up"* (P10).

Finally, much of the recent research focused on sharing wearable fitness data on popular social networking platforms. We asked participants about the impact of this sharing on their behaviors and goals. Our findings contradict those reported by Chung et al. that sharing wearable fitness information on popular social media can encourage physical movements [3]. In our study, the majority of the participants (9/14) who shared their sensed fitness data on common social network sites indicated that this sharing was not all that helpful, and some are no longer sharing on such platforms. Our participants reported several reasons that include lack of interest over time, lack of interest by audience to see this type of information, unclear impact on behaviors and goals, and privacy concerns regarding third party access to their data, especially if the data is shared on Facebook:

*"I don't think it's impacted me that much on Facebook because it is just kind of a general, you know people post things on there and it doesn't have much of impact on me I think"* (P14).

However, the Chung et al. study was based on an existing built-in feature within a Chinese social network for sharing fitness activities. Cultural norms and expectations may explain the difference between their findings and ours. Unfortunately, we did not have the data to examine these factors.

### 4.4 Privacy Concerns

Finally, we explored users' concerns and perspectives regarding privacy of sharing personal and sensed data related to fitness. The overwhelming perception is that most wearable fitness information, and in particular step count, is not sensitive. Thus, few had concerns over sharing this information with any audience:

*"I wouldn't really care if someone knew how many steps I have taken"* (P3).

*"This is not really confidential private information. I mean in some sense it is, but it is not at the level of confidence or privacy that would make think oh I better not to share this"* (P28).

Some of the participants indicated that they would probably be concerned if the disclosed data contains identifiable information, or if the device stores financial information:

*"If it was something from... I don't know, you have to register your ring with your address and you have to have the credit card number in file, so something like that have my personal details that's not fitness related, then I would be concerned."* (P30).

Rather, participants' biggest concerns centered around the ability to manage their self-image and to comply with social norms of sharing. Norms complicate users' decisions to share wearable fitness information in different ways. For example, participants struggled to reconcile the desire to share with their contacts on different platforms (e.g. Facebook) and to conform to what is considered normal to share on those platforms. For instance, P18 commented about sharing his sensed data on Facebook: *"I'm not going to share my blood oxygenation level with friends or in public. That would be ridiculous."* In addition, participants avoid posting too much or too detailed information in order to not bore others: *"My family and friends will kind of get annoyed if I keep sharing constantly"* (P5).

Other participants felt uncomfortable sharing fitness information with the different kinds of contacts they may have on social media platforms, such as professional colleagues. Others worried that friends may perceive sharing fitness achievements as a way of showing off. These concerns led participants to share less on social network sites, and find other platforms for sharing with people with similar goals.

Therefore, maintaining a good self-image was important for interviewees and drove sharing decisions. Users wanted to communicate a positive image regarding their fitness life to other people. Thus, they reported being selective about the information they share, sharing positive achievements for example, rather than sharing generally and automatically. Participants also chose to not disclose information related to eating and sleeping because they think it might potentially impact how they are perceived by others

Some participants (9/30) did express minor concerns over unintended use of their data. This concern was also reported in several prior studies of wearable tracking devices [13, 14]. For example, interviewees were concerned that their health insurance company could get access to their data in the trackers and tie their insurance premiums to their fitness status. Others were concerned that the devices' companies could pass their data to third parties (e.g. sport or drug companies) without their awareness. A few participants also identified that people or organizations could infer personal facts based upon the data shared, and were thus careful about what identifiable information was shared with strangers or organizations.

Finally, there were some concerns related to information security and physical safety. For example, four participants discussed a security breach as a potential risk, resulting in their data being used outside of their intentions. In addition, the GPS feature was a concern reported by four people, indicating that it can be exploited by stalkers:

*"It's pretty much just the location data that bothers me the most. I don't want people knowing where I am in case there is patterns."* (P26).

We asked participants what they do to protect the privacy of their wearable health and fitness data. They reported spending little or no effort on privacy protection, beyond their

choices on what and when to share information. Many of the participants were unaware of their privacy settings; and those few people who were aware about their settings had not changed them since they started to use the service. We asked the participants to go over their profiles and settings if possible, and some discovered that their platform profile was indeed viewable by the public. The remaining participants stated that they changed their controls only once, and that was when they set up the device. Despite this, many participants complained about the lack of options available in the settings to adjust the desired level of privacy.

A few participants reported other ways they protected their information. Two interviewees indicated that they disclose only basic information on their profiles – as little as possible. Another user stated that he put in fake information when he created the account. Six other participants discussed using standard security mechanisms such as authentication. For example, P3 said: *"I have a user name and a password and I just use that in the Garmin to protect my data."*

## 5. Discussion & Implications

Our results reveal a set of common patterns related to users' sharing goals, their chosen audience, and the resulting choices participants make to disclose and manage their sensed information. Our results confirm previous findings of sharing on social media [5, 11], that users are motivated by accountability, advice, and competition when sharing sensed fitness information with other people, in pursuit of their individual fitness and health goals. Participants also reported helping and providing motivation and emotional support to others. An additional goal we have not previously seen is to track and improve health by sharing with physicians. Users expressed a willingness to share if that sharing was helping them meet their health and fitness goals, and reduced sharing if it was found to not help their goals. In other words, participants were consciously making the trade-off to share information for personal health or financial benefits.

Our results also provide useful insights into users' privacy concerns and behaviors. We found that users' practices have little to do with concerns over the sensitivity of the data. Rather, our study suggests that norms and self-presentation are two key concerns that drive users' choices in what information to share, and with whom. Although research suggests that sharing fitness data on popular social network sites is promising to encourage physical activity (e.g. [9]), we argue that site norms can be a barrier and drive users to find other platforms. For example, users sometimes limit the information shared with their friends in order to manage the impressions of the many different contacts they have on social network sites. Many of our participants wanted to communicate a positive image about themselves by sharing only positive fitness achievements. Thus, those who were doing well with their fitness goals found social network sites as a valuable platform to share these achievements with a broad range of friends, but those who struggled with their goals found more support on platforms where they could connect either to friends or strangers with similar goals.

Examining participants' perspectives regarding sensitivity of wearable fitness data reveals that, for the most part, this data is not perceived as sensitive. Users have a common fallacy that there is *"nothing that I need to keep secret."* This perspective toward sensed fitness data has influenced many users to pay little attention to protect their data, even leaving their device platform profiles with default or public privacy settings. Even though some of our participants gave scenarios of how their information could be misused, they felt that the risk was far-fetched. Additionally, the fact that a device company has been in the market for many years has made users trust that their data is safe. However, incidents have demonstrated that wearable fitness data is valuable to criminals [26]. Research has also demonstrated that very sensitive information can be inferred from seemingly innocuous fitness data [22]. While most of our participants did understand that some information could be inferred from their data, they did not express concrete examples of such sensitive information, or feel that they were very susceptible to negative consequences as a result of such inferences.

Our results also suggest that financial incentives are a powerful motive for sharing, and the availability of various wearable fitness devices today has made sharing with different incentive programs (e.g. insurance companies, employers, pharmacies) much easier. This is evidenced by the considerable number of people in our study who disclosed their information to receive discounts or rewards. For a few, the incentive was the primary driver for sharing with such organizations, rather than being in support of a health or fitness goal. However, our findings suggest that concerns over secondary use of data may discourage users from sharing with such programs long term, or as their health or fitness levels decline. Many participants commented that health insurance companies could potentially utilize fitness trackers to adjust coverage plans, although this had not happened yet.

In the light of our findings, we offer several implications for designing wearable trackers that promote sharing and privacy of fitness-related information:

**Design controls and sharing features around common goals and patterns.** As we noted, users' goals for sharing fitness information vary, and this may require sharing different levels of personal information in different ways. Yet

there are a number of common practices depending on those goals, and the associated platforms used. Thus, device platforms could ease this sharing by providing designs centered around these goals and practices. For example, sharing settings could be designed to allow users to have sharing profiles with different audiences. These settings could reflect common data sharing norms, while still allowing for customization of content depending on the goals. In addition, designers should provide visualization mechanisms to help sharers focus on their goals. For example, if a user's sharing goal is to lose weight, a summary of data for this goal such as calories consumed and distance traveled could be visualized in the interface for the intended audience.

**Methods for sharing with physicians.** Individuals who share their wearable fitness data with physicians expressed a strong desire to directly connect their self-tracking data with health providers in some way. Thus, there is currently an unmet need in how to provide full access to, and useful views of information for health providers. For example, sharing settings could be designed to enable users to provide permission to their personal medical provider to access their data using doctors email address, for example. Such support would reduce the burden of this important sharing, facilitate conversation prior to clinical visits, and encourage long term use and tracking of sensed data in support of health goals.

**Awareness of sharing policies.** Generally speaking, users do not consider most of their wearable fitness data to be harmful to them. Yet, they also do not appear to be very aware of device manufacturer's data practices, and how they share their information [27]. Device manufacturers should present their data policies to users on a regular basis (e.g. semiannually), remind users about their choices, and even explain the possible risks to enhance users' awareness. Wearable fitness companies can share users' data with different external parties, such as drug and sport companies. Yet few of our participants expressed concerns over this potential and how data may be shared without their explicit interaction. And while some users acknowledged the possibility of data inference, few expressed any concerns beyond the use of locational data. Additional research is needed to determine what organizational data practices most concern users, and how to increase the awareness over such practices.

**Privacy Nudges.** Many of our participants discussed how their sharing and perceptions had changed over time, as other studies have also pointed out [7, 23]. Thus, designers should seek solutions that are easy to modify over time. Users should be provided with opportunities to reflect on the audiences for their sharing, and how their sharing has changed. For example, as with other forms of social media sharing, nudges could prompt users to reflect on their audience as they share [28]. Nudges could also be designed to remind users of how their information is being shared, and revisit controls over time.

## 6. Conclusion

We conducted a qualitative interview study, investigating the sharing goals, practices, and privacy concerns of 30 users of wearable fitness devices. Our findings reveal that decisions to disclose information to other people and organizations are primarily influenced by the goals people have when sharing with different audiences, and how well different device and sharing platforms can support those goals. Our results highlight the need for more privacy and sharing features centered around these patterns and the sharing norms on various platforms, to support users in their ultimate goals of improving and maintaining their health and fitness.

## 7. Acknowledgments

We thank the anonymous reviewers for their valuable comments. We also thank Zaina Al-Jallad, Wentao Guo, and Safat Siddiqui for helping in the coding process.

## 8. References

[1] Addonizio, G. (2017). The privacy risks surrounding consumer health and fitness apps, associated wearable devices, and HIPAA's limitations.

[2] Apple Took a Commanding Lead In Wearables Fourth Quarter, As Fitbit Slipped. http://fortune.com/2018/03/01/apple-watch-fitbit-wearable-ranking/ Accessed: 2018-10-28.

[3] Chung, C. F., Gorm, N., Shklovski, I. A., & Munson, S. (2017, May). Finding the right fit: understanding health tracking in workplace wellness programs. *In Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 4875-4886). ACM.

[4] Christovich, M. M. (2016). Why Should We Care What Fitbit Shares-A Proposed Statutory Solution to Protect Sensitive Personal Fitness Information. *Hastings Comm. & Ent.* LJ, 38, 91.

[5] Dong, M., Chen, L., & Wang, L. (2018). Investigating the User Behaviors of Sharing Health-and Fitness-Related Information Generated by Mi Band on Weibo. *International Journal of Human–Computer Interaction*, 1-14.

[6] Fritz, T., Huang, E. M., Murphy, G. C., & Zimmermann, T. (2014, April). Persuasive technology in the real world: a study of long-term use of activity sensing devices for fitness. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 487-496). ACM.

[7] Gorm, N., & Shklovski, I. (2016, May). Sharing steps in the workplace: Changing privacy concerns over time. *In proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 4315-4319). ACM.

[8] Gorm, N., & Shklovski, I. (2016, February). Steps, choices and moral accounting: observations from a step-counting campaign in the workplace. *In Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (pp. 148-159). ACM.

[9] Gui, X., Chen, Y., Caldeira, C., Xiao, D., & Chen, Y. (2017, May). When fitness meets social networks: Investigating fitness tracking and social practices on werun. *In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 1647-1659). ACM.

[10] Klasnja, P., Consolvo, S., Choudhury, T., Beckwith, R., & Hightower, J. (2009, May). Exploring privacy concerns about personal sensing. In *International Conference on Pervasive Computing* (pp. 176-183). Springer, Berlin, Heidelberg.

[11] Kreitzberg, D. S. C., Dailey, S. L., Vogt, T. M., Robinson, D., & Zhu, Y. (2016). What is Your Fitness Tracker Communicating?: Exploring Messages and Effects of Wearable Fitness Devices. *Qualitative Research Reports in Communication*, *17*(1), 93-101.

[12] Lidynia, C., Brauner, P., & Ziefle, M. (2017, July). A step in the right direction–understanding privacy concerns and perceived sensitivity of fitness trackers. In *International Conference on Applied Human Factors and Ergonomics* (pp. 42-53). Springer, Cham.

[13] Lowens, B., Motti, V. G., & Caine, K. (2017, August). Wearable privacy: Skeletons in the data closet. In *2017 IEEE International Conference on Healthcare Informatics (ICHI)* (pp. 295-304). IEEE.

[14] Motti, V. G., & Caine, K. (2015, January). Users' privacy concerns about wearables. In *International Conference on Financial Cryptography and Data Security* (pp. 231-244). Springer, Berlin, Heidelberg.

[15] Munson, S. A., & Consolvo, S. (2012, May). Exploring goal-setting, rewards, self-monitoring, and sharing to motivate physical activity. In *2012 6th international conference on pervasive computing technologies for healthcare (pervasivehealth) and workshops* (pp. 25-32). IEEE.

[16] Naeini, P. E., Bhagavatula, S., Habib, H., Degeling, M., Bauer, L., Cranor, L. F., & Sadeh, N. (2017). Privacy expectations and preferences in an IoT world. In *Thirteenth Symposium on Usable Privacy and Security ({SOUPS} 2017)* (pp. 399-412).

[17] Newman, M. W., Lauterbach, D., Munson, S. A., Resnick, P., & Morris, M. E. (2011, March). It's not that i don't have problems, i'm just not putting them on facebook: challenges and opportunities in using online social networks for health. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work* (pp. 341-350). ACM.

[18] New Wearables Forecast from IDC Shows Smartwatches Continuing Their Ascendance While Wristbands Face Flat Growth. https://www.forbes.com/sites/paullamkin/2016/02/17/wearable-tech-market-to-be-worth-34-billion-by-2020/ Accessed: 2018-11-08.

[19] Nissenbaum, H. (2004). Privacy as contextual integrity. *Wash. L. Rev.*, *79*, 119.

[20] Ojala, J. (2013). Personal content in online sports communities: motivations to capture and share personal exercise data. *International Journal of Social and Humanistic Computing 14*, *2*(1-2), 68-85.

[21] Park, K., Weber, I., Cha, M., & Lee, C. (2016, February). Persistent sharing of fitness app status on twitter. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (pp. 184-194). ACM.

[22] Paul, G., & Irvine, J. (2014, September). Privacy implications of wearable health devices. In *Proceedings of the 7th International Conference on Security of Information and Networks* (p. 117). ACM.

[23] Prasad, A., Sorber, J., Stablein, T., Anthony, D., & Kotz, D. (2012, October). Understanding sharing preferences and behavior for mHealth devices. In *Proceedings of the 2012 ACM workshop on Privacy in the electronic society* (pp. 117-128). ACM.

[24] Stragier, J., Evens, T., & Mechant, P. (2015). Broadcast yourself: an exploratory study of sharing physical activity on social networking sites. *Media International Australia*, *155*(1), 120-129.

[25] Teodoro, R., & Naaman, M. (2013, June). Fitter with twitter: Understanding personal health and fitness activity in social media. In *Seventh International AAAI Conference on Weblogs and Social Media*.

[26] Under Armour Says Data Breach Affected about 150 Million MyFitnessPal Accounts. cnbc.com/2018/03/29/under-armour-stock-falls-after-company-admits-data-breach.html. Accessed: 2019-02-19.

[27] Vitak, J., Liao, Y., Kumar, P., Zimmer, M., & Kritikos, K. (2018, March). Privacy attitudes and data valuation among fitness tracker users. In *International Conference on Information* (pp. 229-239). Springer, Cham.

[28] Wang, Y., Leon, P. G., Acquisti, A., Cranor, L. F., Forget, A., & Sadeh, N. (2014, April). A field trial of privacy nudges for facebook. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (pp. 2367-2376). ACM.

[29] Wieneke, A., Lehrer, C., Zeder, R., & Jung, R. (2016, June). Privacy-Related Decision-Making in the Context of Wearable Use. In *PACIS* (p. 67).

[30] Zhao, X., Lampe, C., & Ellison, N. B. (2016, May). The social media ecology: User perceptions, strategies and challenges. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 89-100). ACM.

[31] Zhu, Y., Dailey, S. L., Kreitzberg, D., & Bernhardt, J. (2017). "Social networkout": Connecting social features of wearable fitness trackers with physical exercise. *Journal of health communication*, *22*(12), 974-980.

# Appendix

## A. Screening Survey

**Part 1:**

1. Please select all the wearable devices you own:
   - ☐ Fitbit
   - ☐ Jawbone
   - ☐ Misfit
   - ☐ Polar
   - ☐ Garmin
   - ☐ None
   - ☐ Other, please specify: _____
2. How long have you been using the device(s)?
   - ☐ Less than three months
   - ☐ More than three months
3. Have you ever shared your information recorded by the device(s) with others?
   - ☐ Yes
   - ☐ No

**Part 2:**

4. What is your first name?
   _____
5. What is your age?
   - ☐ 18-20 year
   - ☐ 21-30 year
   - ☐ 31-40 year
   - ☐ 41-50 year
   - ☐ 51-60 year
   - ☐ >60 year
6. What is your email address?
   _____

## B. Interview Questions

**Demographic Questions:**

What is your age?
What is your gender?
What is your level of education?
What ethnicity do you identify with?
What is your current occupation?

**Questions related to the use of the wearable device:**

1. List all the wearable health devices that you own?
2. (If more than one), which do you use most frequently?
3. Have you used any other devices before? Why?
4. What are your goals of using the device?
5. How frequently and when do you use the device?
6. How and when do you look at your data on the device?
7. How has using the device impacted you?
8. What is your overall impression/satisfaction about using the device?

**Sharing preferences and behaviors:**

9. Have you ever shared your information with other people on the device platform?
   If yes:
   a. What information do you share and why?
   b. With whom do you share this information?
   c. Do they also share information with you?
   d. How do you share the information (what context, how frequently, and when)?
   e. How does the sharing impact your behavior and use of the device?
   f. Does sharing your information on the device help you achieve your goals? How?
   g. Has your sharing behavior changed over time?
10. Can you walk me through your profile? Show me what type of settings you have.
11. Did you change the sharing controls in the interface at any point?
    If yes:
    a. Why?
    b. Did you change it for a particular person or a group? Why?
12. Did your choice of sharing recipients affect how you shared the recorded information?
    If yes:
    a. How?
13. Did the sharing controls in the interface allow you to set your sharing preferences easily?
    If no:
    a. Why not?
    b. What changes/omissions/additions would you suggest to make the interface more usable?
    c. If it had been easier to change the privacy preferences, would you have shared differently? How?
14. Have you ever shared your wearable fitness data on popular SNSs, such as Facebook?
    If yes:
    a. Which ones, how, and why?
    b. With whom?
    c. What types of information and how frequently?
    d. Does anyone share such data with you as well?
    e. How does sharing your information on the SNS(s) impact your behavior and use of the device?
    f. Does sharing your information on SNS(s) help you achieve your goals? How?
    g. Do the controls on the SNS help you to share and manage your information?
    h. Do you have any preferences between platforms (i.e. a device platform or an SNS platform) to share your fitness information? Why?
15. Do you have any other way of sharing your information, other than what we have discussed?
16. In general, does the device's platform support your sharing preferences and goals?

**Questions related to privacy:**

17. What are your concerns regarding the privacy of your information on the device?
18. Have these concerns impacted your use of the device?

19. How do you manage your information on the device?
20. To what extent are you comfortable with the existing privacy settings?
21. Do you feel that you are sufficiently protected or do you desire more protections? What kinds of protection do you need?
22. How do you think your information could be misused?
23. Are you worried that your daily activities will be monitored by another person or party when you use the system?
24. Do you have any additional comments about the privacy and sharing of the device data?
25. Would you like to say anything else before we end the interview?

# "I don't own the data": End User Perceptions of Smart Home Device Data Practices and Risks

Madiha Tabassum[1], Tomasz Kosiński[2], Heather Richter Lipford[1]
[1]University of North Carolina at Charlotte, [2]Chalmers University of Technology
{mtabassu, Heather.Lipford}@uncc.edu, tomasz.kosinski@chalmers.se

## Abstract

Smart homes are more connected than ever before, with a variety of commercial devices available. The use of these devices introduces new security and privacy risks in the home, and needs for helping users to understand and mitigate those risks. However, we still know little about how everyday users understand the data practices of smart home devices, and their concerns and behaviors regarding those practices. To bridge this gap, we conducted a semi-structured interview study with 23 smart home users to explore what people think about smart home device data collection, sharing, and usage practices; how that knowledge affects their perceived risks of security and privacy; and the actions they take to resolve those risks. Our results reveal that while people are uncertain about manufacturers' data practices, users' knowledge of their smart home does not strongly influence their threat models and protection behaviors. Instead, users' perceptions and concerns are largely shaped by their experiences in other computing contexts and with organizations. Based on our findings, we provide several recommendations for policymakers, researchers and designers to contribute to users' risk awareness and security and privacy practices in the smart home.

## 1 Introduction

Internet-connected utility devices, called smart home devices, are starting to proliferate throughout households thanks to a growing selection of available devices along with decreasing prices. From lights to thermostats to whole sets of sensors and actuators, users can now enjoy home automation and hands-free control. Yet to provide this functionality, smart home devices greatly expand the types and amount of information about ourselves and our environments that can be collected and shared. The security of our homes is also now becoming reliant on the security of our digital home devices. Thus, with this new domain come new risks to users' security and privacy. And new questions as to how to support users in understanding, reasoning about, and mitigating those risks.

Research on mental models of the Internet has demonstrated that users are uncertain about how their data is collected, shared and stored online, and that users' perceptions often depend on their personal experiences and technical education [12]. Smart homes are even more interconnected, with a wider variety of personal data collected from people's homes. Thus, we seek to examine in greater detail users' perceptions in this more complex environment of the smart home. Our work also builds on previous interview studies, primarily of technically skilled smart home early adopters, examining general privacy and security perceptions [33] and concerns regarding specific data collection entities [34] in the smart home. We focus on users' mental models of the data practices of their smart home devices, and their related privacy and security perceptions.

Specifically, we conducted a drawing exercise and semi-structured interview with 23 participants who have experience living with multiple smart home devices. We focused on recruiting both more technical participants who installed their devices, as well as non-technical users who were not involved in the installation process.We investigated the following research questions: (1) What are end users' mental models of the data flows in their smart home? (2) What are end users' perceptions of the data collection, sharing, storage and use by smart home devices and their manufacturers? (3) How do these mental models and perceptions relate to users' privacy and security concerns, considerations and behavior?

We found that the sophistication of participants' threat model and the adoption of protective measures do not depend on their knowledge of how their smart home works. While participants mentioned some threats and protective measures, they often estimated the privacy and security risks from their smart home devices to be too low to trigger any actions.

Our study makes the following contributions:

- Provides a thorough analysis of both technical and non-technical users' perceptions of smart home device manu-


*USENIX Symposium on Usable Privacy and Security (SOUPS)* 2019.
August 11–13, 2019, Santa Clara, CA, USA.

facturers' data practices and related threats that offer new insights and also confirm, explain, and extend findings from previous studies.

- Among other findings, our results provide new evidence that people are moderately aware of the sensitive information that can be inferred from smart home data, however, are not concerned over the collection and sharing of this data.
- Participants' lower-risk perceptions are shaped by trust, previous experiences within other computing contexts, and their own biases, despite their concerns over the lack of control over their information.

Based on our findings, we provide recommendations for smart home designers, researchers, and policy-makers to provide improved awareness and control of data collection practices and protection strategies, considering the perceptions and capabilities of general smart home users.

## 2 Related Work

User mental models have been explored in the context of the Internet and software. Usually these have been explored from a task- or tool-specific perspective, such as understanding of the operations of WiFi networks [15], general home computer security [31] or firewalls [14, 28]. For example, Kang et al. explored user mental models of the Internet in general [12], also asking users about their perceptions with regard to data practices on the Internet. They found that participants with a more accurate understanding of the Internet identify significantly more privacy threats than participants with simpler models, but that does not influence their protection behaviors. We believe many of these models will carry over into the smart home. Yet, the smart home is more complex, and is more integrated with people's personal lives, introducing new and unique security and privacy risks. Thus, we aim to examine the mental models of smart home users specifically, focusing on the perceptions of data practices of end users.

A number of researchers have examined end user concerns, expectations and preferences with smart devices in the home. However, early work relied on prototypes or probes within homes to examine users' reactions and perspectives, given the limited availability and adoption of smart home devices at the time. For example, Choe et al. [6] used sensor proxies in 11 households as a cultural probe and found participants had concerns about unintended use of their data and the possibility of data exfiltration. They also found tensions between different members of a household around the use and adoption of such in-house sensing applications. Worthy et al. [32] installed an ambiguous Internet of Things(IoT) device in 5 participants' homes for a week and found that trust in the entities that use the data (in this case, the researchers) is a critical factor in the acceptance of the smart device. Montanari et al. [25] invited 16 participants to interact with two smart home devices during the study session and found that users are primarily concerned with the ownership of their data.

A number of studies have also examined the role of context in users' comfort of sharing IoT-related data. These studies reveal that privacy concerns are indeed contextual, depending on a variety of factors such as the type of data recorded, the location where it is recorded, who the data is shared with, the perceived value of the data and benefits provided by services using that data [5, 7, 10, 15, 18, 19, 21, 23, 26]. Naeini et al. [26] used vignettes to study many of these factors with over 380 different use cases across 1,000 users. Their results indicate that people are most uncomfortable when data is collected in their home and prefer to be notified when such collection occurs. Similarly, a survey study by Lee and Kobsa [19] found that monitoring of users' personal spaces, such as their homes, was not acceptable to participants, as well as monitoring performed by the government or unknown entities.

Other studies have found that people are most concerned with certain types of data, namely videos, photos, and biometric information, particularly when this information is gathered inside the home [4, 9, 19, 20, 26]. In another large vignette study, Apthorpe et al. [5] found that participants' acceptance of data collection and sharing was dependent on both the recipient of the information and the specific conditions under which the information was shared. Their results also suggest that users' privacy norms may change with continued use of specific devices. However, results of a different vignette survey by Horne et al. [11] suggest that those changes are not always towards more acceptance of data-sharing. Each of the above studies examines fine-grained contextual factors through survey methods of potential use cases of smart home devices. Despite these findings showing significant concerns over data collection in the home, many users are installing smart home devices that do collect and share such information. These prior studies have not revealed what adopters of current devices think is actually occurring, and their comfort and concerns with those practices.

With widespread adoption, several studies have recently examined the perceptions of users of consumer devices that they use in their own homes and found less concern by actual, regular users. Lau et al. [16, 17] conducted a combination of a diary and interview study with 17 users and 17 non-users of smart voice assistants. They found that the lack of trust and perceived utility are the main reasons for not adopting the device. They also noticed that adopters of the voice assistant have an incomplete understanding of the privacy risks and rarely use existing privacy controls. Most similar to our study, Zeng et al. [33] conducted an interview study of 15, primarily technical, smart home users and observed limited concern among participants about the potential improper use of their data. They also found that even relatively technical participants have an inaccurate or incomplete understanding of smart home technology, resulting in incomplete threat

models and adoption of insufficient mitigation techniques to resolve potential threats. Zheng et al. [34] interviewed 11 technologically skilled smart home users on their reasons for purchasing smart home devices and the perceptions of privacy risks from these devices. They found that users' concerns over specific external entities (i.e. government, manufacturers, internet service providers and advertisers) are influenced by the convenience they get from the device and those entities. While these two interview studies highlight many general concerns of users, and feelings about data being accessed by specific entities, we believe that a more detailed understanding of users' perceptions of the data practices of their smart home devices is critical to understanding users' behaviors and needs. In addition, these prior studies relied primarily on technically knowledgeable participants who actively set up their smart home and are interested in technology, which may limit generalizability of their results.

# 3 Methodology

We conducted a semi-structured interview study and drawing exercise of smart home residents to elicit their mental models of the data practices of smart home devices, along with their perceived security and privacy risks and concerns.

## 3.1 Participants

We sought participants who are regular users of smart home devices and thus had mental models of the smart home ecosystem informed by their usage. We recruited participants with at least three devices, similar to Zheng et. al. [34]. We explicitly recruited some participants who did not install the devices themselves (such as family members) to find people who are not as tech-savvy and may have different privacy perceptions. The participants were recruited through advertisement on Craigslist, and IoT-related Reddit communities. Potential participants were asked to fill in a pre-screening survey answering what types of devices they have in their home, whether they set up the devices by themselves as well as demographic information and email address. We recruited participants until we felt we had a sufficiently diverse sample, and then found we reached saturation (i.e., no new codes or new information attained) during analysis, and hence did not seek additional participants.

We recruited a total of 23 participants (see Appendix A.3). Six of them had a background in computer science, either as a student, or as a computing professional or both. 13 participants were male and six were more than 51 years old. All participants were living in the United States, except one in Canada and one in Sweden. 11 participants installed and manage the devices in their home, 3 participants installed some of the devices and 9 were not involved in the installation and configuration process at all. Not surprisingly, participants who installed their devices self-reported a higher level of familiarity (statistically significant) with technology and smart home security and privacy, than users who did not perform the installation. We acknowledge that there can be tech-saavy non-installers; however, we did not find such participants in our study sample.

## 3.2 Procedure

The researchers contacted selected participants via email to schedule a phone or Whatsapp interview. The interview was semi-structured, with a set of basic questions that were varied depending on the response of the participants. The interviews were recorded via Google voice or an external audio recorder. Interviews lasted on average an hour and participants were given a $10 Amazon gift card for participating. The study was approved by our university Institutional Review Board (IRB).

We started the interview by asking general questions on what smart home devices participants have, and how they use and control those devices. Participants were then instructed to perform a drawing task to elicit their understanding of how their smart home works. Participants were asked to "draw how these devices collect information and how that information flows between the devices and any other involved entities" and to explain their thoughts verbally during the drawing exercise. This has been used as an effective method in capturing mental models in the literature [12, 33]. We utilized remote Google drawing as it was accessible to most of the participants and has been used previously for remote drawing tasks [33]. This could impact the drawings, as the participants utilized shapes and lines rather than free-form strokes. However, participants explained their drawings as they were creating them, similar to an in-person interview. Only 2 participants sent pictures of their drawings via email during the interview because they felt more comfortable drawing on paper. However, after sending the drawing, participants extensively talked about what they drew. We recognize that a drawing exercise in a remote interview is challenging, but we feel the trade-off in finding a more diverse sample was worth it.

We then focused on participants' perceptions of data practices, asking the participants what data they think the smart home devices they own are collecting and where these devices are sending and storing that data. Participants were then prompted to discuss who they think has access to their data and how it is being used, as well as whether the devices are sharing the information, with whom and for what benefit.

Next, we asked participants if they have any concerns regarding those data practices. We then asked them what they do to mitigate their concerns and resolve the threats that they think arise from using their smart home devices. We discussed what controls the participants believe they currently have over the data the devices are collecting, what controls they expect to have and their expectations regarding the security of their data. Finally, we collected participants' demographic infor-

| Type of device | Count | Examples | Users' perception of information collection |
|---|---|---|---|
| Intelligent voice assistant | 20 | Google Home, Amazon Echo | Voice interaction (20); Usage (10); Account info (5) |
| Smart light | 16 | Philips hue, LIFX, Sangled | Patterns & usage (11); State of the lights (10); |
| Smart plug and switch | 13 | Wemo, Tplink, Insteon, Sonoff | Account info (5); Home location (2) |
| Smart camera/doorbell | 11 | Nest Cam, Ring, SkyBell Doorbell | Video (11); Home location (4); Usage (3) |
| Smart thermostat | 11 | Nest, Ecobee Thermostat | Temperature (10); Usage (5); Energy use (3); Account info (2) |
| Hardware hub | 8 | Samsung SmartThings, Wink hub | Usage (6), Location (3), Other devices in the network (2) |
| Streaming device | 8 | Roku, Fire Sticks, Chromecast | Viewing history (4); Account info (3) |
| Other devices: Smart TV (5), Leak sensor (4), Smart Doorlock (3), Open/close sensor (3), Motion sensor (3), Smoke detector (2), Smart media hub (2) | | | |

Table 1: Summary of the devices owned by participants. Numbers in the parentheses are number of participants

mation at the end of the interview. Interview questions are provided in Appendix A.2.

### 3.3 Data Analysis

We transcribed the interviews and used an inductive coding process to analyze the data. Two researchers independently coded the interviews of five participants and came up with a list of common themes and patterns. Then the researchers compared and merged the themes and agreed on a shared codebook with 15 structural codes divided into 60 sub-codes. The two coders then independently coded the rest of the interviews. After all the interviews were coded, the researchers met and discussed the codes, resolving any disagreements caused by misunderstanding the codes. We tracked the disagreements and the Cohen's Kappa, a measure of inter-rater reliability, was calculated at 96.37.

The participants' drawings and related verbal explanations were separately analyzed by the primary author, who clustered similar drawings and conceptions into two emerging categories. The clustering was performed based on the complexity of participants' mental model about both the physical architecture of their smart home and corresponding data flows throughout the system. These categories were then discussed among all the authors, and used to examine differences between participants' perceptions throughout the results.

### 3.4 Limitations

As with similar interview-based studies, we consider sample size to be the biggest limitation of this work. We can only provide limited qualitative results on the posed research questions, yet hope that those revealed patterns can be used in formulating further studies of more representative populations and to inform design. We also believe that the participants, even the non-technical ones, that we interviewed are still the early adopters. They are clearly well educated, and likely of high socio-economic status. They also value the benefit of the devices and decided to have them in their homes. Hence, they have already made the decision that the trade-off is worth the risk; therefore they may not have as many concerns as non-adopters. Thus, these results may not generalize to a broader consumer base who will adopt smart home devices in the future. Still, we hope that many of these patterns would be found in a more general population as we found many of the perceptions did not differ between participants of different levels of expertise. Another limitation is that this was a one time interview, which entails the risk of missing participant concerns that could be discovered in, for instance, a longitudinal study. Finally, almost all of our participants are from the U.S. and may have a different perspective about privacy from other regions. Because we have only two participants from other countries, it was not enough to identify those differences.

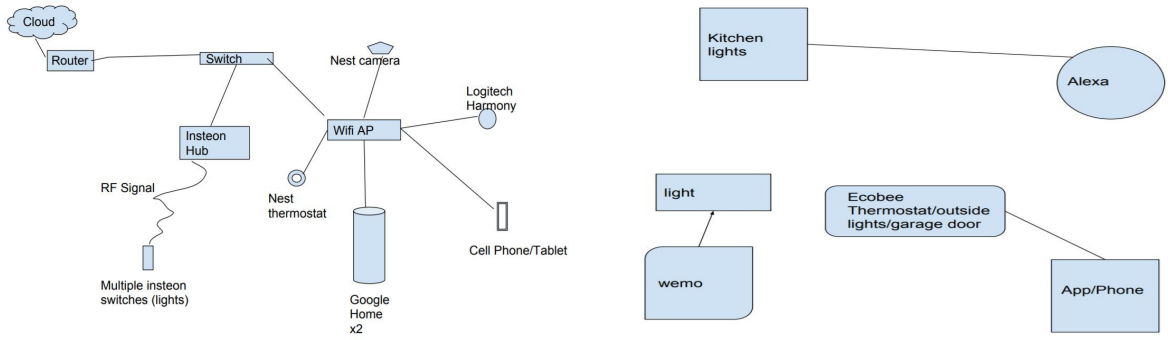
## 4 Results

Our study goals are to examine users' perceptions and concerns of the data practices of smart home devices. First we describe the devices they have and use, then present the results of our analysis of participants' mental models, their perception of manufacturers' data practices and their related security and privacy concerns and behaviors. Please note that the numbers reported below are not meant to convey quantitative results, but simply reflect the prevalence of particular themes within our experimental sample.

### 4.1 General Use of Smart Home Devices

Participants own a wide variety of internet-connected devices, including integrated devices (lights, thermostats), home monitoring and safety devices (security cameras, door locks), home appliances (vacuum cleaners, smart refrigerator), and intelligent personal assistants (Google Home, Amazon Echo). We summarize the common devices in Table 1. Participants use these devices in a number of ways. The most frequently mentioned ($n = 11$) use case is household automation (automatically turn on/off the lights, adjust the temperature, etc.), followed by remotely sensing and controlling the home ($n = 10$) (i.e. to turn on/off the lights, check on pets). Another use case ($n = 9$) is increasing the security or safety of the house (by notifications of conspicuous sounds in the house, water leakage, etc.). Other less frequently mentioned use cases are energy saving and help with household chores.

We also asked participants how they interact with their devices. Participants use several different methods, often in combination, depending on the location of the user within or outside of the home, as well as the type of device and




(a) Advanced model (Id9)

(b) Service-oriented model (Id20)

Figure 1: Drawing of participants with different mental models

its compatibility with a controller. Almost all participants ($n = 21$) have a central controller set up, i.e. either a smart voice assistant, hardware hub, an app (e.g. Apple Homekit) or a custom-made controller using the Raspberry Pi. For 13 participants, voice is the primary method of interaction when they are home, utilizing either Amazon Echo or Google Home. Some participants also mentioned setting up triggers based on other sensor data or timers to make devices fully automated (i.e. using IFTTT services).

## 4.2 Mental Models of Smart Home

Our analysis shows that participants with different technical backgrounds and experiences with the devices have different mental models of how their smart home works. We asked users to describe how data flows in their smart home, and participants chose different ways to express this. We grouped based on similarity of participants' understanding of how devices are connected and how information flows in the smart home and this resulted in our categorization. Two models emerged: advanced (9 participants) and service-oriented (12 participants), based on participants' drawings and verbal explanations of their smart home. We did not include Id6 and Id21 in our categorization. The recording of Id6's drawing explanation was distorted, and Id21's spouse was helping her with the drawing during the interview.

Participants with the advanced model consider their smart home as a complex, multi-layer system. These participants have a reasonable understanding of the logical topology of the smart home, connection mechanisms (Ethernet, WiFi, ZigBee/Z-Wave) and the role of some network components (routers and hub) in communication (Figure 1a). All the participants with this model also discussed how data flows back and forth between the devices and servers in the cloud when interaction happens. For example, Id19 said,

"*When its (Echo) not being used, it is just waiting for one of four trigger words and when that triggers, then it opens up the connection back home(Amazon) and start parsing out the commands for different devices and passes it along to the smart things which takes over from there.*"

Participants with the advanced model discussed how information flows through the infrastructure as well as to the companies' servers and comes back to the device. These participants personally installed all of their smart home devices. Also, a number of them did some customization in their smart home, i.e., used IFTTT to automate the devices, installed a personal server or built a central assistant using Raspberry Pi. It can be one of the reasons behind their more comprehensive understanding of the network topology. Additionally, these participants are also more informed of the complexity of the flows as well as the fact that devices are sending information to companies' servers as soon as they interact with them.

Participants with the service-oriented model (n=12) have a reasonable understanding about which devices communicate with each other inside the house, but do not have deeper technical knowledge of how that communication happens other than via the WiFi. Their mental models of the smart home mostly consist of the interaction between the smart devices (i.e. lights) and the controller (e.g. Google Home) they use to control the device, but no awareness of the role of other networked components in the device interaction (Figure 1b). There were a few participants in this group who brought up that information is going to the cloud initially when drawing their smart home; the other participants didn't. However, when asked directly during later interview questions they all indicated that information the devices are collecting is not stored locally, it is leaving their home to the cloud or some server. However, the participants with the service-oriented model expressed no or very shallow awareness of the role of the cloud in the device interaction.

## 4.3 End Users Perception of Data Practices

In our analysis, we found that participants' mental models of how their smart home devices work do not often relate to their perceptions of smart home device data collection, usage, and sharing practices. Rather, their understanding is primarily based on interactions they have had with the devices

or what they see in the corresponding applications. Some participants ($n = 10$) beliefs of data practices were informed by their perception of particular companies and experiences with those companies in other (non-smart home) contexts, which sometimes leads to inaccurate conclusions on what really happens. For example, Id11 thinks that companies will not sell any data because it would upset their consumers and the companies would lose reputation. Only two explicitly reported privacy policies as a source of their knowledge about data practices. Below we discuss the findings on end users' perceptions of data practices in detail.

#### 4.3.1 Data Collection:

Not surprisingly, participants' perception of data collection was informed by the type of the device and their experience with that device. For example, all participants who have a smart doorbell or camera were aware of the device collecting video recordings, but none demonstrated any awareness of the corresponding applications tracking their location whenever they use it. In other words, participants were well aware of the primary data that the device is collecting but may overlook secondary data that does not directly correlate with the type of device or basic utilities received from the device. In Table 1, we summarize user perceptions of what information is collected for different devices. Only the devices owned by more than five people are listed.

For most of the devices, participants believe that their usage and interaction patterns are being recorded and they were not that concerned about the data collected by the smart home devices. We also looked more specifically at audio and video data since previous studies [26, 34] found that these data are more sensitive to people. When put in practice, video is still considered as the most sensitive; however, participants for the most part were able to find practices that allowed them to be comfortable with the collection of video. For example, using the camera in a live streaming mode without recording the video, starting video recording only when the house is empty, or using an outdoor camera or video doorbell, so nothing inside the house gets recorded. For instance, Id7 mentioned:

"*Only information I would potentially ever be concerned with, like the way I use my device, is the images on the camera. But again, the camera is turned off when I am home, and on when I am not home.*"

However, in one extreme case, a participant (Id22) removed an indoor camera from her apartment. She reported being unable to use the camera outside because of concern of residents of her apartment complex. She was not aware of other alternative configurations for her camera such as using the camera only for live streaming or removing the recordings from the cloud, which she may have been more comfortable with.

Participants did not show much concern about the collection of their audio data. They know that the voice assistant is recording after the trigger word, and they were comfortable with the audio being recorded in that way. One non-installer participant (Id20) was uncomfortable with the voice assistant as she suspected that Amazon Echo may be listening to her even if she is not calling it using the trigger word. Her Echo had recently showed an Amazon package delivery notification with yellow lights, and she misinterpreted it as the device listening to her conversation. Even though her husband later clarified the misunderstanding, the participant was still very uncomfortable and did not want the device in her house at the time the interview was conducted. Another technical participant decided not to buy any commercially available voice assistants because of a worry over companies harvesting the audio.

Despite their awareness, many participants ($n = 15$) believe smart home devices are collecting more information than they should. However, some participants ($n = 9$) said the data collection was mostly positive. These participants explicitly mentioned that most of the data these devices collect is needed in order to either provide them the services they expect or to make the devices more convenient to use.

We also asked participants what can be inferred about them from the data these devices are collecting. In contrast to a previous study [34], we found that participants are somewhat aware of the sensitive information that can be inferred from the seemingly innocuous data collected by the smart home devices. For instance Id11 mentioned,

"*They can probably tell that I don't have the lights everywhere at my home. That I am out of the house during the day time. They can probably tell when I am sleeping because the lights are not turned on that time.*"

The types of inferred information that were mentioned are: habits and preferences (i.e. buying habits, music preferences, etc.; 14 participants), daily schedule (i.e. when home or not, when using which devices, etc.; 11 participants), tentative location of the house (8 participants), other occupants in the house (i.e. have pets or kids; 3 participants), political views (2 participants), sleeping patterns (2 participants) and other devices in the house (2 participants). Three explicitly mentioned that these companies can infer a lot of things that consumers can't even imagine.

#### 4.3.2 Data Storage:

When prompted, all participants reported being aware that at least some of the smart home device data is being stored externally, with twenty specifically mentioning the cloud or a server operated or owned by the manufacturer of the device. However, three other service-oriented participants expressed a vague idea such as 'somewhere in some kind of database.' For example, Id15 said:

"*I don't know really where it goes or what happened to it but I imagine that it does get stored somewhere, some kind of database and somebody is able to analyze and see different trends through it. But I have no idea.*"

Eleven participants explicitly mentioned there is either no or very limited local storage of the data, that everything is stored in the cloud. Participants frequently mentioned they have no control over the data once they shared it; however, some ($n = 5$) hypothesized that it might be possible to remove their data by contacting the device manufacturers. Interestingly, 4 participants suspected that even if they remove the data, it will still be in the cloud. A number of participants ($n = 8$) also mentioned companies are doing the bare minimum to protect their consumers' data in the server. Most participants were not sure about companies' data retention practices except for the retention period of the video. Some of them also made interesting inferences, for example, five participants believe that Google and Amazon store data forever or for a very long time because these companies have enough resources to store such data, while smaller companies do not.

Interestingly, all the participants who installed the camera or video doorbell themselves ($n = 7$) know about the video deletion option or after how many days the video will be automatically removed from the server. On the other hand, participants who have not installed ($n = 3$) the camera or the video doorbell are not sure about the storage policy of the video or the option of deleting the video. Video is the one exception where some participants are very aware of the data storage practices and available controls, but only those who installed it, and as a result, they found practices that they were comfortable with and configured their device accordingly. But, participants who are not the installer did not get that understanding, which in one case led to a lot of discomfort and removal of the device.

We did not find as many difference between installers and non-installers regarding their knowledge of data storage policies and controls provided by the devices that collect audio. Out of 20 participants who had a smart voice assistant, 15 are familiar with the device usage log where they can review their voice interactions with the assistant. However, some of them either are not familiar with the data deletion option ($n = 5$) or skeptical that Amazon or Google may keep the data even after they delete it from the log using the available interface ($n = 4$). However, all the participants who did not know about the device usage log were also not involved in device installation. For one participant, this lack of awareness also lead to more discomfort about using the device, as stated by Id20:

"*I have asked my husband to disconnect the Alexa(Echo) multiple times. Just because I'm not comfortable with it. But if it did collect data, I would have no idea how to find it and to remove it so I would just disconnect it.*"

#### 4.3.3 Data Use:

Participants discussed three primary uses of the data their smart devices collect. The most frequently mentioned use case is targeted advertising or marketing to sell products to consumers ($n = 19$). For instance, Id19 said:

"*They have put a lot of money in this product, and then they are selling it. So, they must be using it for something other than me telling my house to turn on my bedroom light. They are building advertising model of me. They want to know who I am and how I work so they can try to sell me something.*"

Participants were aware that their habits, preferences, and daily schedules can be inferred from the data smart devices are collecting and can be used for targeted advertising. However, targeted advertising seems to have became so integral to participants' lives that they accepted it as a price of living in the age of the Internet.

Many ($n = 17$) mentioned that the companies are using the data to improve the current product, for instance by fixing malfunctions/errors (4 participants), improving the user experience or tailoring the device to customers needs (4 participants) or improving the services provided by the device (2 participants). As Id7 stated:

"*(Companies use the information) in order to better the products I guess. I guess if there are errors like you know if I ask Google Home to do something, and the lights don't respond, they're surely collecting that kind of information*"

A number of participants ($n = 9$) also believe that the information companies are gathering can help them to recognize users' needs and come up with new products.

#### 4.3.4 Data Sharing:

Participants identified a number of entities that they believe have access to the data their smart home devices are collecting: the manufacturer of the device/the data analysts working with the company ($n = 23$); third parties/advertisers interested in the data ($n = 9$); parent companies, subsidiaries or affiliates of the device manufacturers ($n = 7$); hackers ($n = 7$); legal organizations such as government security agencies ($n = 4$); the manufacturer of the device/app that is used to control the device ($n = 3$) and other people who have accounts with the device ($n = 2$).

We then asked participants if they think companies share any information with third parties. Twenty-two participants agreed that they do. Nine further believe that companies are sharing only their demographics or preferences but not any personal information; however, 4 participants mentioned they believe companies are sharing everything. Participants also made interesting inferences about how the sharing happens, such as that the big companies (Google, Amazon, Apple) do not share data at all while only the small companies share their consumers' data (6 participants). For example, Id8 said:

"*I think Amazon would be like the top consumer of this information; I think they're collecting this for themselves. I don't think they would share it. I think a smaller company... if the Ring wasn't purchased by Amazon, I think Ring might share that information with Amazon...I have a feeling that's why Amazon bought them.*"

Most of the participants ($n = 18$) said they agreed to this

sharing by signing the terms of service or privacy policy or saying 'yes' to everything during the installation process. But similar to previous research [13, 22], participants reported not reading privacy policies and pointed out the usability issues of such agreements. Three service-oriented participants believe they consented just by using the product. Some participants ($n = 9$) stated that once the data is sent to the cloud, it is out of their hands and control. Id12 stated:

*"I'm sure they do... absolutely they do it (share data)... they are allowed to do that...they can do whatever they want with it, that data is considered as their property. They can keep everything for their own or they share."*

Many participants reported that the only way they can opt out from this sharing is to stop using the product ($n = 15$), while a few mentioned modifying the applications' settings for partial opt-out ($n = 4$) or by contacting the company (n=2).

To summarize participants' perceptions of data practices: they base their understanding of what data is collected on their experiences and interaction with the devices. For the most part, they expect that their data resides in the cloud and that it can be and is shared by companies, with little ability to control that. However, participants expressed a great deal of uncertainty when they discussed the ways companies are collecting, using and sharing their data. The only exception is the video data where all the participants who installed the device were aware of where the video is stored and video retention time. Several participants ($n = 5$) explicitly expressed their concern about companies not being transparent enough about their data practices. Many participants mentioned that they want more transparency from the device companies ($n = 14$). For instance, Id9 said:

"*If these companies are sharing my data with third parties, I'd like to know who they are sharing with, maybe like if I go to the Insteon website they say, hey we share your data here. So a website that keeps track of all this stuff would be good.*

Participants also want companies to take enough measures to ensure their data is protected ($n = 9$). A few participants ($n = 4$) also believe there are not enough regulations in place and that policymakers should enact and enforce more strict laws to protect consumer data. Finally, ten participants expressed the desire to have explicit control over data collection and sharing and to be able to remove their data from the cloud.

## 4.4 Security and Privacy Threats and Consequences

We now turn to participants' perceptions of the risks and behaviors for protecting their information. Participants identified several threats and discussed how these affect their security and privacy. However, we again could not find many differences between participants with different technical knowledge levels and mental models. Instead, many of the concerns participants mentioned came from their experiences with the Internet, computers and mobile phones instead of threats specific to smart home devices.

### 4.4.1 Threats:

The most concrete and frequent threat mentioned by participants ($n = 17$) is a data breach in the cloud and their personal information being compromised. Two participants also suggested hackers could gain access to aggregated profile data from the cloud. Id2 stated his concern as:

"*I mean especially the states of data breaches lately. That is concerning because they're not viewing in a way that hey, these are actual consumers out there, these are real people. Then they may not have the best security practices, and that data can get out somewhere.*"

Some participants ($n = 11$) also pointed out that their smart home devices or the WiFi can be hacked and remotely controlled by adversaries for various reasons, i.e. to spy on them, break into their house, etc. For example, Id19 said:

"*someone could access my lights, someone could turn my heat up ... umm ... if I had a smart lock, someone could have access that to get in my house but I don't have a smart lock. Just like I wouldn't use banking through any of these devices because the consequences are too severe in case there was a breach... the same with a lock, I wouldn't use one of those.*"

Six participants also identified improper use and sharing of their data with third party companies as a potential threat. Unlike data breaches and device hacking, participants were more vague about this threat, i.e., third party companies may use my data for some nefarious reasons or their server may not be secure, etc. Id12 said:

"*The person you shared that data with can share the data with somebody else. Like if you shared data with the company that follows all the rules and if they share with a company that doesn't follow any rule that is out there. I don't think these companies have any methodologies in place to ensure that whether their partner will maintain the data safety or not.*"

### 4.4.2 Consequences of the threats:

These threats were then associated with specific negative outcomes. Similar to the concerns expressed in previous papers on smart homes [33, 35], participants most frequently mentioned the violation of their physical security and safety ($n = 10$). They implied that smart home devices know when they are home or not, and what other devices they have in their home, and that this information can be used to rob them or physically harm them. Id3 mentioned:

"*I guess if it was a criminal group like a gang or something they could use that data to know when I'm home or not home. If they want to rob, what is the best time to rob, where to go in my house, what my house looks like, that kind of information.*"

Participants also mentioned the possibilities of identity or financial theft ($n = 4$). Three advanced participants expressed

their discomfort about the abilities of companies to manipulate their decisions, judgment or perception of things in some way. Id23 said: "*I think they can show me what I like; I think they can alter the world I am living into the world that is preferential to me, as a consumer.*"

Other risks that participants identified are profiling (n=2), criminals/companies using data to uniquely identify people (n=2), spear phishing (n=1) and social engineering (n=1).

Interestingly, some participants ($n = 6$) shared a general discomfort around the feeling of surveillance, of people knowing too much information about them and being able to use that for nefarious reasons specially around the devices that collect audio and video. For instance Id20 mentioned:

"*Makes me feel uncomfortable that I am in my own home and I can't just say whatever I want without somebody listening you know?*"

Participants with the advanced model identified more examples of threats, and 8 of the 9 were concerned with data breaches. However we found no additional differences between participants based on their mental models. In line with the previous work [33], we found that despite participants identification of these threats, only a few expressed significant concerns or worry about them. However, participants did take some actions to protect the security and privacy of their smart home as we will further discuss below.

## 4.5 Protective Measures

Participants reported a diverse range of protective measures that they perform or are aware of to reduce their security and privacy risks. Both traditional security best practices and use of protection tools/services were discussed by participants.

### 4.5.1 Behavioral/non-technical mitigations

Many participants ($n = 12$) mentioned self-censoring their way of using smart home devices. It took various forms, such as turning the device off, changing behavior in front of the device, or avoiding the use of certain device functionality ($n = 6$), as well as limiting the amount of information disclosed to the device ($n = 8$) by not providing more information than absolutely necessary while signing up for an account, or by using someone else's account. For instance, Id22 mentioned changing her behavior in front of the camera:

"*It knew when I woke up and walked to the kitchen... it is in the living room... so it kind of sees that I come around the corner to the kitchen...I kind of try to stay by the wall because I didn't want my robe or pajamas or whatever I was wearing to be on camera.*"

Some participants ($n = 8$) also expressed concerns about their financial information and mentioned frequently monitoring their bank accounts and using credit monitoring services.

### 4.5.2 Technical mitigations:

Participants discussed using various traditional technical security practices ($n = 9$), such as changing and using strong passwords and using two-factor authentication. Two also reported using certain devices offline to limit access to their data. Two participants with the advanced model also discussed using a separate network for smart home devices. Id8 stated:

"*I have a closed WiFi network for my IoT devices. I do password changes and what not, also my WiFi isn't broadcasted.*"

### 4.5.3 Tool-based mitigations:

Participants also discussed using some tools or services to protect their privacy around smart home devices ($n = 7$). Two participants hosted local servers and customized the devices to work with that. Others mentioned using different network security devices, installing firewalls or a VPN to protect their network from outside attacks. Id3 stated:

"*I do have a firewall set up on my network that apparently helps with if people try to get the data from me... I can't do anything about the data stored on the cloud. Hopefully the firewall cuts down on any devices that might be compromised or part of a botnet or something like that.*"

A number of participants ($n = 5$) expressed their awareness of such tools or services but were not using those at the time the interview was conducted.

The tool-based mitigations were primarily discussed by the more technically knowledgeable users; nine of the twelve who mentioned tool-based mitigations had the advanced mental model. Furthermore, only the participants with advanced mental models demonstrated familiarity with customizable tools/services for preventing their data from being sent out to the Internet ($n = 5$). On the contrary, most of the participants with the service-oriented model attempt to mitigate their concerns by following traditional security practices (e.g. changing passwords) derived from other computing contexts or changing their behaviors around the devices.

In summary, participants have demonstrated an understanding of some risks from the smart home, but they are not very concerned about many of them. Only a few technical participants did use tools specifically to protect their smart home. Others kept on following the best practices they know from other contexts either because they don't know about what actions to take in the smart home context or the cost of finding and taking those actions is way bigger than their concern. Participants discussed a number of reasons for their lack of concern and unwillingness to take protective measures, as discussed in the next section.

## 4.6 Reasons for lack of concern and protective actions

While participants could all discuss perceived threats to their security and privacy, most did not express strong concerns.

Several themes emerged when we asked participants why they are not concerned about their security and privacy in the smart home.

**Acceptance of trade-off:** Most of the participants (n=15) mentioned that they have to give up some of their data and accept the risks for the convenience and services provided by these smart home devices. Four participants also mentioned feeling powerless over this trade-off. For instance, Id12 said:

"*Once I bought all these devices that was it. These functions come with these risks no matter what and I can't do anything about that. There are no third option. If you want the device you have to accept those risks, otherwise don't use it at all.*"

Though participants accepted the trade-off between their privacy and the convenience, 13 of them stated a desire for more transparency from the device manufacturers.

**Trust of the manufacturers:** Another common reason was participants' trust in the device manufacturers. Eleven participants stated that they trust that companies will not misuse their data because it would damage the company's reputation or will not be financially profitable. Id7 said,

"*I don't think they (companies) are selling it to Russian, I don't think they are trying to steal my identity. I don't think there's anything other than just trying to improve the product, trying to use the information for marketing and advertising.*"

**Optimism bias:** A number of participants ($n = 9$) expressed a low likelihood of being affected under the assumption that they are not an attractive target for hackers. For instance, Id10 mentioned: "*I also went to college and have student debt. So, I don't feel like an attractive target for someone to try to steal my identity or really do anything.*

**Marginal risk:** Participants tend to judge the risk from smart home devices by comparing it with how exposed they already are. Several participants ($n = 9$) were not concerned because they believe a wide array of information about them has already been collected or available otherwise and the smart device won't increase the risk. For instance, Id13 said:

"*I've been using the Internet since like I was in middle school... so I don't really have an expectation of privacy.*"

Ten participants believe the data that smart devices are collecting are not that useful or sensitive and would not be harmful to them in the future. Five participants also explicitly mentioned not being concerned because smart devices do not have any critical information about them, i.e., financial details, SSN, etc. Id16 mentioned:

"*I would be worried about just the things like my credit card information or maybe like social security... that hasn't been shared with any other companies... as for like my habit I don't really think that's (concerning) because the companies will only be able to tailor the things we want.*"

Three of these participants also felt that they have already taken enough action to keep their smart home safe.

**Trust of regulators:** Four participants believe that there are appropriate regulations or overseeing bodies in place which will protect their data from potential misuse by companies. Id19 said: "*If they(company) violate it(rules) it's either going to be corrected or will be most likely to be shut down by a government agency or something.*"

**High cost of protective actions:** A few participants (n=3) with the advanced mental model also discussed the inconvenience of implementing useful protective measures. For example, Id9 explained the inconvenience of locally hosting the services:

"*You know if I wanted some services that did not connect to the Internet then I kind of have to purchase that myself and run everything that way to prevent, you know, things on my network from going out to the Internet.*"

## 5 Discussion

We will now report the key insights learned from our study and discuss implications and recommendations for designers, policy makers and researchers.

**Knowledge of smart home does not influence threat model or trigger actions:** Even though participants had different levels of understanding about how their smart home works, their perception of device manufacturers' data practices was quite similar and not much different from the findings of the earlier work on Internet perceptions [12]. Furthermore, our participants' knowledge about their smart home and manufacturers' data practices did not affect their awareness of possible threats in the smart home. Rather, participants with advanced and simple mental models both frequently mentioned threats and protective actions that are known from the context of the Internet, but also applicable in the smart home. However, participants with the more advanced mental model did show more awareness of the protective measures unique to the smart home, such as preventing data from going outside of the home. Yet, despite awareness of the threats and protective measures, most of the participants choose not to put those into practice. Instead, participants' decisions of protective actions were more influenced by their own biases and concerns related to general Internet usage.

**Difference in knowledge (or a lack thereof) between different participant groups**: The two groups that emerged in our analysis, i.e., participants with the advanced and service-oriented model, seem to differ primarily in their technical detail and understanding of their smart home. While the participants with advanced model were all installers, there were installers with the service-oriented model as

well. However, we did not find many differences between participants with these two mental models and installers vs. non-installers in terms of their perceptions of data practices. The only difference in knowledge is that the installers of smart cameras and doorbells are more aware of companies' video data storage practices. One reason for installers having this awareness can be the fact that the users need to buy an additional subscription to store the video in the cloud for many of the devices (i.e., nest aware subscription for nest camera, ring protect plan for ring doorbell ). This added step exposed the installers to the company's policy regarding video data storage.

Users' lack of exposure to companies' data related policies, in general, may be the reason for the similar perceptions of different groups of participants. This asserts the need for including such information about data practices as a part of the application that is used to control the device and designing nudges and cues for users (installers and non-installers) to get exposed to that information.

**Trust paradox**: Participants know about much of the data collection occurring with their smart home devices. Many of them are also aware of companies' lack of security in the cloud and data sharing with third-party organizations. Some of them also believe that there is not enough legal protections for consumers. Yet, participants justified their lack of concerns and protective actions with trust that companies will not misuse their data as it will tear down their reputation and regulators will close the company. This paradox can be explained by the notion of learned helplessness seen in many participants, where they ignored possible negative consequences because they feel they have no control. Participants described how once data is collected from their devices, it's beyond their control. And sometimes coped by censoring themselves in some way to keep data from being captured by a device and entered into an application in the first place. Participants thus primarily rely on the organization to keep their data secure and expect governments and policymakers to regulate what is occurring, rather than taking many actions by themselves.

**Estimated risk is too low to take action:** One of the main reasons for inaction is that participant's estimated risk from the smart home devices is quite low. They are aware of the fact that their daily schedule and habits can be inferred from the data smart home devices are collecting and that companies may use that for targeted advertising. However, companies have been using data such as buying habits for targeted advertising for a long time; it was nothing new to the participants and not viewed as an added risk. Even the risk of a break-in was also not able to raise participants' concerns as they believed they would not be a potential target. A number of participants also didn't think that the use of smart home devices may increase their risk of identity theft as they think there is already enough information out there on the Internet if someone wants to target them specifically. Even the participants who have been a victim of identity theft were quite comfortable with their smart devices as they believe they put enough protection on their financial accounts. None of the participants showed awareness about news of potential smart home device or data misuse, and may not realize the breadth of risk imposed by their devices. Rather, all the participants accepted the trade-off between the benefit of smart home devices with their lower perceived risk as mentioned by Id19, "*I wouldn't let something that I personally see so small affect something that I am enjoying using so much. Something that I personally think more serious, like access to my bank and things like that.. I would lock it down and stop using it immediately.*"

**Lack of awareness about data practices and controls impede usage:** Despite participants' perceptions and expectations of a large amount of data collection and sharing, we also note that participants are still very uncertain about the device manufacturers' data practices, echoing prior work on users' perception of the Internet and cloud storage more generally [3, 8, 12]. Many participants were also uncertain or unaware of the controls they have on their devices. For a few participants, these uncertainties led to not using certain device functionalities or using the device only at specific times or specific places and may also influence their freedom of expression. In two extreme cases of non-installer participants, Id20 and Id22, it led to the desire of removing the device from their house. However, from their interviews, it appeared the awareness of the available controls may have influenced their privacy behaviors, as mentioned by Id22, "*If I had an easy way to do it... if I had to push a button to remove it(camera recordings) then I would surely remove it.*" In other words, more familiarity with controls may have led those participants to be more comfortable using the device. This underscores the importance of future research to examine ways to nudge users, especially those who are not involved in the set-up and configuration of their smart home, to discover and utilize the available controls.

## 5.1 Implications and Recommendations

**Enhance transparency and control:** People want more transparency and control over the data collected and shared by smart home device manufacturers. Participants should have the ability to remove the data and set sharing preferences of their data where possible, for instance, sharing only aggregated data, sharing only usage data, etc. Companies can provide more transparency and controls to users by designing a dedicated web-page or privacy setting in the mobile application where users can view the data points collected by the devices. Another suggestion is to provide privacy and data-related information in addition to the set-up information

in the box, which as Peppet [27] reported, many of the IoT device manufacturers do not. Multiple participants appreciated Google for the transparency and added control in their devices, whereas some were more skeptical about buying devices from lesser-known companies. New smart home start-ups can improve their reputation by providing more transparency and control over users' data.

Researchers have also proposed and developed dedicated devices and tools to give users more security and privacy controls [2, 29, 30]. For instance, Karmann et al. developed 'Alias,' a device that paralyzes the voice assistant by preventing it from listening and only activates the assistant for a custom wake word from the user [2]. Mennicken et al. proposed a calendar-based interface, Casalendar, that visualizes triggered actions and the sensor data collected in a smart home to facilitate users' understanding [24]. We advocate for more such research on novel security and privacy tools and controls beyond the features currently available within a device. While few of our participants were actively looking for additional tools, we believe that easy to use off-the-shelf tools, if commercially available, may increase the comfort of privacy-sensitive people and provide more options for privacy preserving use and adoption.

**Best practices for companies and users:** As smart home devices become more widespread, smart home attacks will also become more common. Yet, participants who have simpler mental models of their smart home are often aware of and adopted only common traditional best practices (i.e. changing passwords) that may not always help against the security and privacy risks unique to the smart home. Current measures that can help (i.e. locally hosted services) are too technical for the vast majority of potential users. Yet, it is also unclear what best practices are - what are the best methods for average consumers to protect themselves, their data, and their homes? Thus, we concur with Zeng et al. [33] that security researchers, policy makers, and manufacturers need to develop an additional set of best practices for smart home users. However, we want to emphasize that such best practices should be developed by keeping the mental models of users and their technical capabilities in mind. Our findings also revealed that participants rely on companies and policy makers to protect their data. With the widespread use of multiple smart home devices, it will be burdensome for users to manage and take responsibility for all of the data collected and shared by smart home devices. Our study also reinforces the need for the enforcement of a set of privacy best practices for smart home device manufacturers [34]. Policymakers should consider how to administer these rules and penalize companies that do not comply with regulations.

**Develop mechanisms to increase user awareness about visual indicators and controls:** Researchers need to explore how additional awareness mechanisms can be incorporated directly into smart home devices and applications. For instance, exploring ways to nudge users toward available controls or designing observable cues that provide added awareness of data collection and sharing. For example, Amazon Echo shows blue light patterns when it starts listening. However, designers need to be careful while designing visual indicators, as we found that use of similar indicators (i.e., showing yellow light patterns as a delivery notification by Echo) can be confusing to users. In addition to developing visual indicators, designers should also explore ways to inform users, especially non-installers, of those indicators as a primary part of interaction with the device. For instance, on the first interaction with new users, the voice assistant can speak out loud about the controls they have over their data.

**Educate people about future risk:** Most of the recent news on IoT misuse is about the use of devices for Distributed Daniel of Service attacks. People do not feel personally targeted when they learn about such generalized attacks. Furthermore, even though participants were aware of the sensitive information that can be inferred from their smart home data, they were unaware of how that data can be used other than for advertising. Centralized online resources are needed where people will be able to learn about the data practices and possible risks from different smart home devices, so that existing users can asses their risk, and potential buyers can decide whether and which device to buy. Mozilla already provides one such online guide [1], however none of our participants mentioned it. Strategies should be taken to educate users about possible risks and available public resources to find information about their devices.

## 6 Conclusions

In this qualitative interview study of smart home users, we found that participants generally understand that a wide range of information is being collected about their interactions with smart home devices, and shared with a variety of entities to provide useful functionality as well as for marketing and advertising. Much of this information is stored in the cloud, where it is out of the control of users. Yet users are also highly uncertain about these data practices, and desire greater awareness and control over what is occurring. Participants also identified several threats common across computing contexts - such as breaches and financial theft, as well as home safety and security. Yet, despite this awareness of potential threats, they did not view these as serious risks and practiced few mitigation strategies beyond trying to provide devices with no more information than necessary. These findings provide new information about how users perceive what is occurring in the smart home and suggest the need for greater awareness and user friendly control mechanisms as well as cues and visual indicators to inform and contribute to users' security and privacy practices in their homes.

## Acknowledgments

We thank our user study participants and pilot participants for their time and input. Tomasz Kosiński was partially supported by the Wallenberg Artificial Intelligence, Autonomous Systems and Software Program (WASP), funded by the Knut and Alice Wallenberg Foundation.

## References

[1] Mozilla - *privacy not included. https://foundation.mozilla.org/en/privacynotincluded/. Accessed: 2019-02-13.

[2] Project alias. http://bjoernkarmann.dk/project_alias. Accessed: 2019-02-13.

[3] Mark S. Ackerman, Lorrie Faith Cranor, and Joseph Reagle. Privacy in e-commerce: Examining user scenarios and privacy preferences. In *Proceedings of the 1st ACM Conference on Electronic Commerce*, EC '99, pages 1–8, New York, NY, USA, 1999. ACM.

[4] Noura Aleisa and Karen Renaud. Yes, I know this IoT device might invade my privacy, but I love it anyway! a study of Saudi Arabian perceptions. In *2nd International Conference on Internet of Things: Big Data and Security (IoTBDS 2017)*, pages 198–205, 2017.

[5] Noah Apthorpe, Yan Shvartzshnaider, Arunesh Mathur, Dillon Reisman, and Nick Feamster. Discovering smart home internet of things privacy norms using contextual integrity. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(2):59:1–59:23, July 2018.

[6] Eun Kyoung Choe, Sunny Consolvo, Jaeyeon Jung, Beverly Harrison, Shwetak N. Patel, and Julie A. Kientz. Investigating receptiveness to sensing and inference in the home using sensor proxies. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, UbiComp '12, pages 61–70, New York, NY, USA, 2012. ACM.

[7] Eun Kyoung Choe, Sunny Consolvo, Jaeyeon Jung, Beverly L. Harrison, and Julie A. Kientz. Living in a glass house: a survey of private moments in the home. In *UbiComp*, 2011.

[8] Jason W. Clark, Peter Snyder, Damon McCoy, and Chris Kanich. "i saw images i didn't even know i had": Understanding user perceptions of cloud storage privacy. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 1641–1644, New York, NY, USA, 2015. ACM.

[9] Anupam Das, Martin Degeling, Xiaoyou Wang, Junjue Wang, Norman Sadeh, and Mahadev Satyanarayanan. Assisting users in a world full of cameras: A privacy-aware infrastructure for computer vision applications. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 1387–1396. IEEE, July 2017.

[10] Marco Ghiglieri, Melanie Volkamer, and Karen Renaud. Exploring consumers' attitudes of smart TV related privacy risks. In Theo Tryfonas, editor, *Proceedings of the 5th International Conference on Human Aspects of Information Security, Privacy, and Trust (HAS)*, Lecture Notes in Computer Science, pages 656–674, Cham, 2017. Springer.

[11] Christine Horne, Brice Darras, Elyse Bean, Anurag Srivastava, and Scott Frickel. Privacy, technology, and norms: The case of smart meters. *Social Science Research*, 51:64 – 76, 2015.

[12] Ruogu Kang, Laura Dabbish, Nathaniel Fruchter, and Sara Kiesler. "my data just goes everywhere:" user mental models of the internet and implications for privacy and security. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*, pages 39–52, Ottawa, 2015. USENIX Association.

[13] Z. Kaupas and J. Ceponis. End-user license agreement-threat to information security: a real life experiment. In *Proceedings of the IVUS International Conference on Information Technology*, pages 55–60, 2017.

[14] Predrag Klasnja, Sunny Consolvo, Tanzeem Choudhury, Richard Beckwith, and Jeffrey Hightower. Exploring privacy concerns about personal sensing. In *Proceedings of the Seventh International Conference on Pervasive Computing*, 2009.

[15] Predrag Klasnja, Sunny Consolvo, Jaeyeon Jung, Benjamin M. Greenstein, Louis LeGrand, Pauline Powledge, and David Wetherall. "when i am on wi-fi, i am fearless": Privacy concerns & practices in eeryday wi-fi use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 1993–2002, New York, NY, USA, 2009. ACM.

[16] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. "alexa, stop recording": Mismatches between smart speaker privacy controls and user needs. https://www.usenix.org/sites/default/files/soups2018posters-lau.pdf. Accessed: 2018-09-10.

[17] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. Alexa, are you listening?: Privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW):102:1–102:31, November 2018.

[18] Scott Lederer, Jennifer Mankoff, and Anind K. Dey. Who wants to know what when? privacy preference determinants in ubiquitous computing. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '03, pages 724–725, New York, NY, USA, 2003. ACM.

[19] H. Lee and A. Kobsa. Understanding user privacy in internet of things environments. In *2016 IEEE 3rd World Forum on Internet of Things (WF-IoT)*, pages 407–412, Dec 2016.

[20] Linda Lee, Joong Hwa Lee, Serge Egelman, and David Wagner. Information disclosure concerns in the age of wearable computing. In *Proceedings of the NDSS Workshop on Usable Security (USEC '16)*. Internet Society, 2016.

[21] Nathan Malkin, Julia Bernd, Maritza Johnson, and Serge Egelman. "what can't data be used for?" privacy expectations about smart tvs in the us. In *Proceedings of the 3rd European Workshop on Usable Security (EuroUSEC)*.

[22] Thomas Maronick. Do consumers read terms of service agreements when installing software?a two-study empirical analysis. *International Journal of Business and Social Research*, 4(6), 2014.

[23] Faith McCreary, Alexandra Zafiroglu, and Heather Patterson. The contextual complexity of privacy in smart homes and smart buildings. In *HCI in Business, Government, and Organizations: Information Systems*, pages 67–78, Cham, 2016. Springer International Publishing.

[24] Sarah Mennicken, David Kim, and Elaine May Huang. Integrating the smart home into the digital calendar. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 5958–5969, New York, NY, USA, 2016. ACM.

[25] Alessandro Montanari, Afra Mashhadi, Akhil Mathur, and Fahim Kawsar. Understanding the privacy design space for personal connected objects. In *Proceedings of the 30th British Human Computer Interaction Conference (British HCI 2016)*, 07 2016.

[26] Pardis Emami Naeini, Sruti Bhagavatula, Hana Habib, Martin Degeling, Lujo Bauer, Lorrie Faith Cranor, and Norman Sadeh. Privacy expectations and preferences in an iot world. In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*, pages 399–412, Santa Clara, CA, 2017. USENIX Association.

[27] Scott R. Peppet. Regulating the internet of things: First steps toward managing discrimination, privacy, security, and consent. *Texas Law Review*, 93:85–179, 11 2014.

[28] Fahimeh Raja, Kirstie Hawkey, and Konstantin Beznosov. Revealing hidden context: Improving mental models of personal firewall users. In *Proceedings of the Symposium On Usable Privacy and Security (SOUPS)*, 01 2009.

[29] A. K. Simpson, F. Roesner, and T. Kohno. Securing vulnerable home iot devices with an in-hub security manager. In *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pages 551–556, March 2017.

[30] V. Sivaraman, H. H. Gharakheili, A. Vishwanath, R. Boreli, and O. Mehani. Network-level security and privacy control for smart-home iot devices. In *2015 IEEE 11th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pages 163–167, Oct 2015.

[31] Rick Wash. Folk models of home computer security. In *Proceedings of the Sixth Symposium on Usable Privacy and Security*, SOUPS '10, pages 11:1–11:16, New York, NY, USA, 2010. ACM.

[32] Peter Worthy, Ben Matthews, and Stephen Viller. Trust me: Doubts and concerns living with the Internet of Things. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS '16)*, pages 427–434, New York, 2016. ACM.

[33] Eric Zeng, Shrirang Mare, and Franziska Roesner. End user security and privacy concerns with smart homes. In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*, pages 65–80, Santa Clara, CA, 2017. USENIX Association.

[34] Serena Zheng, Noah Apthorpe, Marshini Chetty, and Nick Feamster. User perceptions of smart home iot privacy. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW):200:1–200:20, November 2018.

[35] Verena Zimmermann, Merve Bennighof, Miriam Edel, Oliver Hofmann, Judith Jung, and Melina von Wick. 'home, smart home' - exploring end users' mental models of smart homes. In Raimund Dachselt and Gerhard Weber, editors, *Mensch and Computer 2018 - Workshopband*, Bonn, 2018. Gesellschaft für Informatik e.V.

# A Appendix

## A.1 Recruitment Survey

- How many smart home devices do you own?
- Please select all the smart home devices you own? (Choices: A list of devices with option to include devices that are not listed)

- Do you have a degree on any Computer Science related major? (Choices: Yes, No)
- Who installed and automated the smart home devices in your house? (Choices: I installed all the devices, I installed some of the devices, Someone else installed all the devices )
- Your name .....
- Your email .....
- Your Age (Choices: Less than 20 yrs, 21-30 yrs, 31-40 yrs, 41-50 yrs, 51-60 yrs, More than 60 yrs)

## A.2 Interview Questions

**General questions:**

- What smart devices do you have in your house?
- How did you use these devices?
- How do you control these devices?

**Drawing exercise:**

- Can you draw how these devices collect information and how that information flows between the devices and any other involved entities?

**Data collection: (for each device)**

- What information is collected by the device?
- Do you think that data should be collected?
- Do you think it needs to be collected? If so, for what purpose?

**Data storage**

- Where do you think the device transmits this information?
- Where do you think the data are stored? What data are stored? For how long?
- Is it possible to check what data are stored? If yes, how?
- Do you have any control over the stored data?
- Is it possible to remove this data? Have you ever considered removing data?
- Can you remove your device usage log?

**Data sharing:**

- Who can access and use the data that have been stored?
- How do the device manufacturer/others use the data?
- Does your device manufacturer share these data with any other companies and organizations? If yes with whom?
- Why do you think they share the data? What are the benefits? To them and to you?
- Do you think you opt-in to this sharing? When and how do you opt into sharing?
- Do you think you can opt out? Do you consider opting-out?
- Are the devices sharing data between themselves? What data, how and for what purposes?

**Data inference:**

- Does that concern you about the way the device manufacturers use your data? What are some of the concerns?
- How can a third party use your data? Does that concern you? What are some of the concerns you have regarding this?
- What can be inferred about you from this data by the entities or organizations that have the data?
- What do you think some of the threats are to your data or yourself?

**Mitigation techniques:**

- Have you done anything to resolve these threats and to protect your data?
- What you think you should be doing?
- What controls do you have on your data? How hard it is to use these controls?
- What controls do you want to have or would like to be able to do regarding your data privacy?
- What do you think companies are doing to protect your data privacy? What do you expect them to do?

**Closing question:**

- Is there anything else or any concern you want to share with me about your smart home or expected me to ask?

**Demographics:**

- What is your ethnicity?
- what is your primary occupation?
- What is the highest level of education you have completed?

- What was your major?
- Did you have any degree on a computer science related topic?

**Self-reported technical skill [33]:**
On a scale of 1(very weak) - 5(very strong)

- How would you rate your knowledge of technology in general?
- How would you rate your knowledge of computer security and privacy?
- How would you rate your knowledge of smart home technology?

## A.3 Summary of Participants' Demographics

| ID | Gender | Age | Education | Profession | Installed the devices? |
|---|---|---|---|---|---|
| ID1 | M | 21-30 | MS: Computer Engineering | Grad student | Yes |
| ID2 | M | 21-30 | BS: Computer Science | Programming consultant | Yes |
| ID3 | M | 21-30 | Juries Doctorate | Attorney | Yes |
| ID4 | M | 31-40 | Doctorate: Medicine | Product manager | Yes |
| ID5 | F | 21-30 | BS: Biology | Banking | No |
| ID6 | M | 61-70 | BA: Urban Planning | Retired computing professional | Yes |
| ID7 | M | 51-60 | Associate Degree: Arts and Science | Computing professional | Yes |
| ID8 | M | 41-50 | Diploma: Media Arts | Network engineer | Yes |
| ID9 | M | 31-40 | BS: computer science | IT sales | Yes |
| ID10 | F | 31-40 | MS: Kinestheology | Unemployed | No |
| ID11 | F | 21-30 | MS: Kinestheology | Clinical researcher | Yes |
| ID12 | M | 31-40 | Post Graduate: Chemistry and Physics | Business entrepreneur | Yes |
| ID13 | F | 31-40 | MS: educational counseling | Education administration | Yes |
| ID14 | M | 51-60 | BA: Criminal Justice | Banking | No |
| ID15 | F | 31-40 | BA: Russian | Human Resource | No |
| ID16 | F | 21-30 | Bachelors: Biology and Psychology | Insurance verification specialist | No |
| ID17 | M | 31-40 | Masters: Sociology and Applied Research | Higher education administrator | Yes |
| ID18 | F | 21-30 | Bachelors: Elementary Education | Fifth grade teacher | No |
| ID19 | M | 31-40 | High School | Customer Service | Yes |
| ID20 | F | 61-70 | Bachelors: Accounting | Accountant | No |
| ID21 | F | 61-70 | College | Retired | No |
| ID22 | F | 51-60 | BA: Practical Civilization | Administrator: call center | No |
| ID23 | M | 21-30 | BS: Biomedical Sciences | Graduate student | Yes |

# More than Smart Speakers: Security and Privacy Perceptions of Smart Home Personal Assistants

Noura Abdi
*Department of Informatics*
*King's College London*
*United Kingdom*
noura.abdi@kcl.ac.uk

Kopo M. Ramokapane
*Bristol Cyber Security Group*
*University of Bristol*
*United Kingdom*
marvin.ramokapane@bristol.ac.uk

Jose M. Such
*Department of Informatics*
*King's College London*
*United Kingdom*
jose.such@kcl.ac.uk

## Abstract

Smart Home Personal Assistants (SPA) such as Amazon Echo/Alexa and Google Home/Assistant have made our daily routines much more convenient, allowing us to complete tasks quickly and efficiently using natural language. It is believed that around 10% of consumers around the world already own an SPA, and predictions are that ownership will keep rising. It is therefore paramount to make SPA secure and privacy-preserving. Despite the growing research on SPA security and privacy, little is known about users' security and privacy perceptions concerning SPA complex ecosystem, which involves several elements and stakeholders. To explore this, we considered the main four use case scenarios with distinctive architectural elements and stakeholders involved: using built-in skills, third-party skills, managing other smart devices, and shopping, through semi-structured interviews with SPA users. Using a grounded theory approach, we found that users have incomplete mental models of SPA, leading to different perceptions of where data is being stored, processed, and shared. Users' understanding of the SPA ecosystem is often limited to their household and the SPA vendor at most, even when using third-party skills or managing other smart home devices. This leads to incomplete threat models (few threat agents and types of attacks) and non-technical coping strategies they implement to protect themselves. We also found that users are not making the most of the shopping capabilities of SPA due to security and privacy concerns; and while users perceive SPA as intelligent and capable of learning, they would not like SPA learning everything about them. Based on these findings, we discuss design recommendations.


*USENIX Symposium on Usable Privacy and Security (SOUPS)* 2019.
August 11–13, 2019, Santa Clara, CA, USA.

## 1 Introduction

The adoption of smart home personal assistants (SPA) has rapidly increased in the last few years [5]. Estimates suggest that 10% of the world consumers own an SPA [37], and that over 50 million Amazon Echo devices have been sold to date in the US alone [27]. SPA benefit from recent advances in Natural Language Processing to handle a wide range of commands and questions in a playful way, with a name and a gender assigned to the SPA, which encourages users to personify them and therefore interact with them in a human-like manner and be more engaging [32]. SPA are used to shop, stream music, and set timers, alarms and reminders among many others [43].

Despite the numerous benefits and convenience SPA bring, they also raise security and privacy concerns. Prior work, including [12, 17, 22, 28], already highlighted numerous security and privacy issues in general with smart home technologies and in particular with SPA. In addition, very recent research also studied users' privacy concerns with SPA [19, 30], but this research typically centred around privacy and the smart speaker part of the SPA ecosystem. However, smart speakers are just the tip of the iceberg, i.e., an SPA is normally composed of at least a smart speaker such as Amazon Echo *and* a cloud-based voice assistant such as Amazon Alexa. Also, the SPA ecosystem is complex and includes several parties: the SPA provider, multiple third-party providers of skills or actions that SPA can request following users' voice commands (e.g. playing music through Spotify), and multiple providers of other smart home devices (e.g. smart bulbs) being managed through the SPA.

To bridge this gap, in this paper we focus on the following research questions. What are users' perceptions of the SPA architecture and the SPA data ecosystem? What threat models do users have concerning SPA? What mitigation strategies do users use to alleviate risk and other challenges they face?

To answer these research questions, we conducted semi-structured interviews with seventeen current SPA users. Following a grounded theory approach, we interviewed people

who had been using Amazon Echo/Alexa and Google Home-/Assistant, which are the two most used SPA and together dominated circa 87% of the SPA market as of 2017 [39]. We particularly asked about their use of the SPA, how they think SPA process and complete their requests, as well as other data activities like storage, sharing and learning using the four main use cases of SPA: built-in skills (such as setting reminders and alarms), third-party skills (such as Spotify and Uber), managing other smart devices (such as smart bulbs and smart TVs), and shopping. We also elicited users' threat models and the strategies they use to protect themselves when using SPA.

Our contributions include:

- We present users' understanding of SPA's ecosystem, discussing their conceptions and misconceptions about how data is processed, stored, shared and learned by SPA and the actors involved through four main use cases of SPA (built-in skills, third-party skills, managing other smart devices, and shopping). We show that users have a limited understanding of SPA, which leaves them with very inaccurate and at best incomplete mental models of the SPA ecosystem.
- We uncover the lack of trust users have with some of the use cases of SPA, in particular shopping, and how this is hampering adoption of these use cases, providing the reasons we found behind this phenomenon.
- We report the threat models users have of SPA, showing both threat agents and types of attacks users consider possible. We also show the mainly non-technical coping strategies users follow to try to protect themselves.
- and, we present design implications for how SPA might support users' expectations and needs with regards to privacy and security.

## 2 Background

Smart Home Personal Assistants (SPAs) have a complex architecture [14], as depicted in Figure 1, that usually involves at least a smart speaker (e.g. Amazon Echo, Google Home) and a cloud-based voice personal assistant (e.g. Alexa, Google Assistant). A normal request works as follows, the user utters a request to the smart speaker, which is then processed in the SPA provider's cloud using Natural Language Processing to understand users' speech and intent. Once the intent is identified, the SPA provider delegates the user request to a set of *Skills*[1]. Skills provide users with functions such as the ability to play music, check weather updates, control other smart home devices and shopping. There are currently over 70,000 Alexa skills [1] and 2,000 Google Assistant skills [34]. There are two main types of Skills: Built-in Skills provided by the SPA provider (e.g. Weather updates, Shopping) and Third-party Skills provided by third party developers using the development Skill Kits (e.g. Spotify, Smart Home Devices). Importantly, third-party skills are typically hosted in a remote web service host controlled by the developer of the third-party skill. Finally, any outputs produced by a Skill are sent back to the SPA Provider, which generates a spoken response, which is then push backed to the smart speaker, which plays the response to the user.

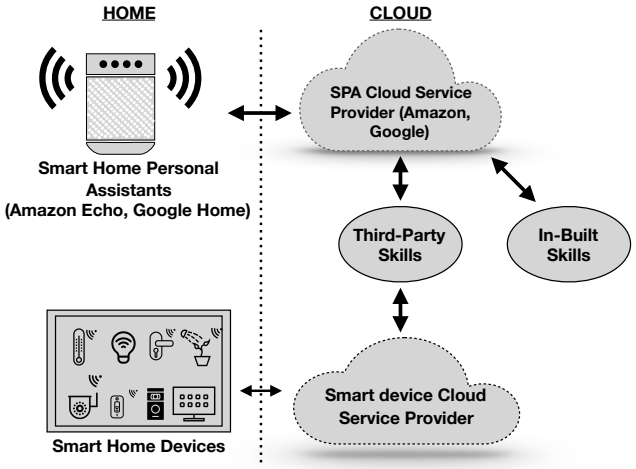



Figure 1: Sketch of the SPA Ecosystem (inspired by [14]).

## 3 Related Work

In this section, we first discuss research conducted on users' security and privacy perceptions in the Smart Home in general. Then, we discuss research that focused on the security and privacy of SPA in particular.

### 3.1 Security and Privacy of the Smart Home

Extensive research has been conducted on the security and privacy of smart homes. For example, from a more human factors point of view, prior work studied users' mental models for smart home devices. Zeng et al. [46] conducted semi-structured interviews on fifteen smart home owners examining users' mental models about their device. They found that users who had advanced mental models about their device were those with highly technical level of understanding regarding their smart home system whilst those with intermediate level of mental models showed some level of understanding on how their smart home system works [46]. Similarly Zheng et al. [47] conducted semi-structured interviews with eleven smart home owners to be able to understand their privacy

[1]Note that, for easy of exposition, we adopt Amazon's terminology of Skills, but these may be called differently in other SPA platforms. For instance, in Google Assistant, skills are called *Actions* instead.

perceptions of the devices. Their work highlighted that smart home owners prioritize convenience over privacy and will allow their data to be shared if their perceived benefits outweigh privacy risks. Also users perceived that it is the device manufactures responsibility to protect users privacy. More recently, Emami-Naeini et al. [15] studied security and privacy perceptions of IoT device owners examining their concerns prior and after purchase. Users were asked to rank important factors when they are considering to purchase an IoT device, with security and privacy ranking highly important. They also showed users a security and privacy label prototype aimed at helping them make better security and privacy decisions when purchasing IoT devices [15]. In [36], the authors studied smart home security identifying issues that influence or affect security decisions in the home, e.g., perceived competence, trust and cost were some of the factors identified. Finally, He et al. [24] examined access control specification and authentication in the home IoT, looking at different access controls that can be applied for different tasks depending on the context. While the works above considered the smart home, including SPA, they did not consider the SPA ecosystem in full.

### 3.2 Security and Privacy of SPA

There has been an increasing amount of research focusing exclusively on SPA security and privacy. One line of research focused on technical attacks and defences. For instance, Haack et al. [22] and Kumar et al. [28] reported vulnerabilities of Amazon Alexa, focusing on the speech recognition ability of SPA (e.g., interpretation errors of user commands exposing the device to outside attacks). In terms of defences, Lei et al. [31] implemented a Virtual Security Button (VSB) which detects the presence of human motion and then prevents unauthorized access. Huan et al. [16] proposed a continuous voice authentication mechanism for SPAs that aims to ensure SPA works solely on commands from a legitimate user. Kepuska et al. [26] proposed a multi-model dialogue system that combines various factors such as; voice, video, head and body movements for secure SPA authentication.

Another line of research focused on human factors of security and privacy in SPA. In particular, previous work studied users' perceptions, including Frutcher and Liccardi [19], who examined users' online reviews of SPA to understand privacy and security concerns. More recently, Lau et al. [30] studied users and non-users reasons for and against adopting SPA. Their findings highlighted that many non-users did not see the benefit in using SPA while users shared privacy risks such as the device listening but would rather trade privacy for convenience. Our work differs from previous work on users' perceptions of SPA security and privacy, as we consider the whole SPA ecosystem, while previous works tended to focus more on the smart speaker part of the SPA only.

## 4 Methodology

To answer our research questions, we conducted a qualitative study following a semi-structured approach [6] and Grounded Theory [8, 20]. We used a pre-screening process and semi-structured interviews as detailed below. The study was reviewed and approved by King's College London IRB.

### 4.1 Pilot Study

We created an initial version of the interview script to explore users' perceptions around our main research questions. Before running the full study, we conducted five preliminary interviews. We recruited interviewees internally within our university with the aim of ensuring that the interview questions were easy for interviewees to understand, did not take too long to complete, and would provide insights with regards to our research questions without guiding or biasing the interviewees. With these aims in mind, we conducted and analyzed the preliminary interviews and refined the interview script twice. None of the data collected during the pilot study was used in the final data analysis.

### 4.2 Recruitment and Screening

We recruited potential participants through Prolific (www.prolific.ac) and internally within King's College London. All potential participants were asked to fill out a screening survey which queried for their demographic information (age, gender, education background, employment status), the SPA and other type of smart devices they own, what they use the SPA for, and how long they have been using the SPA — see Appendix A for the screening questions. The questionnaire took on average 10 minutes to complete and the participants who completed the survey through Prolific were compensated with an average of £1.20.

The screening responses were used to select Amazon Echo or Google Home owners who had been using their SPA for at least one month and had used the device for various tasks such as setting the alarm or reminders, using third-party skills, shopping or managing other smart home devices. Our demographics data also helped us to select participants in a way in which we would maximize demographic coverage. This was done to ensure that selected participants had experience in using the SPA since we wanted to elicit their mental models regarding how SPA work while making sure we had a balanced sample of demographic data. In some cases, the decision was to take everyone who completed the questionnaire in a logical way, but with a particular characteristic, e.g., we invited all valid participants who said they used the device for shopping because of the low number of participants saying they used the device for shopping. Finally, we included some questions designed to rule out participants just pretending to be SPA owners.

The recruitment phase took place between November 2018 and January 2019. The qualified participants were contacted and invited for an interview. Participants were asked to provide their Skype ID. Because this is personally identifiable data, we needed approval from prolific to use such information to recruit participants. We contacted Prolific informing them about our research and the type of data we would be collecting, and our request was approved.

## 4.3 Participants

From the recruitment and screening process, we received a total of 43 (31 prolific, 9 internal) responses, from which 31 qualified for an interview following the criteria explained above. We contacted all of them, and from the 23 who responded to be available for an interview, we then ordered and prioritized them in order to maximize demographics and SPA usage, until saturation was reached — more details about the methodology and data analysis below. In total, we interviewed 17 participants (13 Prolific and 4 internally). Table 1 summarizes demographic and SPA usage information for all the participants. The interviews were conducted via Skype or in person between January 2019 and February 2019, and participants were rewarded with £10 for completing the interview.

Table 1: Summary of Participants.

| | **# participants** |
|---|---|
| **Gender** | |
| Male | 8 |
| Female | 9 |
| **Age** | |
| 18 - 20 | 2 |
| 21 - 25 | 4 |
| 26 - 30 | 5 |
| 31 - 40 | 3 |
| 41 - 50 | 2 |
| 51+ | 1 |
| **Highest Level of Education** | |
| High school/College course | 6 |
| Undergraduate | 2 |
| Graduate | 8 |
| Postgraduate | 1 |
| **Employment status** | |
| Full time | 9 |
| Part-time | 3 |
| Unemployed | 1 |
| Retired | 1 |
| Student | 3 |
| **Device Type** | |
| Amazon Echo | 10 |
| Google Home | 7 |
| **Period of usage** | |
| 1-6 months | 5 |
| 6-12 months | 6 |
| 1-2 years | 4 |
| 2+ years | 2 |
| **Device use** | |
| Set alarm, reminders and checking the weather | 13 |
| Third-party Skills (e.g. Spotify) | 11 |
| Managing smarthome devices | 6 |
| Shopping | 4 |

## 4.4 Interview Protocol

Interviews were led and conducted by the lead researcher. Before the interviews, we provided the participants with an information sheet, which explained the purpose of the study. During the interview, the lead researcher introduced themselves and further explained the purpose of the study. We then asked for consent to participate and record audio.

To make participants feel at ease and establish rapport, the interview started with general questions about the participant and their device, we asked them what type of device they owned, what they use it for, how often they used it, how long they have been using it, and whom they were using it with.

The second set of questions focused on asking participants about other smart home devices they own; what devices they owned and if they use their smart assistant to control or communicate with those devices. Then, we asked participants about how they registered and set up their devices. This included questions about voice recognition and purchasing.

To understand and elicit users' mental models about the infrastructure and the data ecosystem, we created four scenarios regarding how the device is used. We would then ask about each scenario depending on the previous answers of the participants, i.e., if they said they had other smart devices they connected the SPA to, then we would ask about the scenario about managing smart home devices. Each scenario was structured as follows. At the beginning of each scenario, we asked participants to think and describe how the SPA worked to complete each request. The second set of questions asked about data storage (including the requests themselves), where data is stored and for how long. The last set of questions focused on whether data was shared and with whom. We describe the scenarios below:

**Scenario 1 - Built-in Skills**

> In scenario 1, we asked users to think about instances when they asked the device to give them a weather or traffic update. We then asked them to describe how their devices processed their request when using in-built skills. After this, we asked them if such requests are stored, and if yes, where they are stored and for how long. The last set of questions focuses on understanding if data are being used for other purposes than responding to their requests and by whom.

**Scenario 2 - Third-party Skills**

In Scenario 2, we asked users about third-party skills they used (e.g. Spotify) by asking them to describe how they think the process works regarding how requests are processed and handled. We then asked them whether they think their requests are stored and if so where and for how long. Regarding data sharing, we also asked them if they think their data is shared with third-party skill providers as well as other third-parties such as advertisers.

**Scenario 3 - Managing smart home devices**

The purpose of scenario 3 was to understand how users perceived SPA's interaction with other smart home devices — e.g., smart bulbs. We asked users to describe how the SPA controls or manages other smart home devices. We began by asking users to think about instances when they controlled other smart home devices with their SPA. Then, we asked them to describe the process to us. We followed asking them if requests are stored. If the user thought these requests were stored, then we continued to ask them where they were stored and for how long. Regarding data sharing, we first asked users if requests were shared with the provider of the smart home device. Then, we asked if SPA's provider (Amazon or Google) together with the smart device provider shared data with other third-party companies.

**Scenario 4 - Shopping**

In the last scenario, we asked participants to describe how they use the device to shop and how do they think the process works. Similar to other scenarios, we asked them if the device stored their requests including purchase history and for how long. We also asked them if the data was used for other purposes and shared with other third-party companies.

The last set of questions focused on understanding users' threat models concerning the device. Instead of asking participants plainly whether they had security or privacy concerns about using the device, we asked participants what their thoughts were of the SPA capabilities to learn about them based on their interactions with it, who might want to take advantage or exploit the SPA and how, and if they had any concerns about the SPA. Before we concluded the interview, we asked participants about how they protected their devices or mitigated concerns if they mentioned some exploits or other concerns. We provide the final interview script in Appendix B.

## 4.5 Data Analysis

Following a grounded theory approach [8,20], two researchers independently started the coding process immediately after the first two interviews. Coding was started early to identify interesting codes and categories that could be explored in-depth. The interview scripts were then analyzed through several iterative stages of open, axial and selective coding. When new codes or themes emerged, both researchers met and discussed the new findings and amended the interview script where necessary to explore the new codes or themes in depth. Examples of the codes that emerged very early were: useful, best fit, control, and convenient were prevalent. We discussed and coded these codes under "Useful" as a theme that we defined to denote that users found the device to be useful. New codes stopped emerging after the ninth and tenth transcripts, but we stopped interviewing new participants after number seventeen to confirm we had reached saturation, i.e., to check new codes or themes would not emerge. During the selective coding phase, we ordered and grouped our themes into more broad and abstract groupings to answer our research questions.

# 5 Findings

This section presents the results of our study. It is structured as follows. It begins by reporting the results in terms of how users use and setup the SPA and the different parts of the ecosystem. Then, we report the different perceptions users have of data processing, storage, sharing and learning across the SPA ecosystem. After this, we focus on the results about one particular use case: shopping, as we found a general lack of trust in SPA shopping capabilities that we study more in-depth considering users who do not shop at all, users who only do part of the shopping process (e.g. shopping lists), and users who do purchase using the SPA. Finally, we report on the threat models users have and the kind of defences and coping strategies they put in place to tackle the threats.

## 5.1 Device Usage

Participants used SPA for various tasks, all of them falling into the four main use case scenarios:

***Built-in Skills.*** Participants mentioned they used their SPA to complete everyday tasks such as setting an alarm, setting reminders and checking the weather.

***Third-party Skills.*** When asked about third-party skills, participants mentioned using Spotify to listen to music, Uber to call a taxi, Fuel Finder for checking fuel prices, etc.

***Managing Other Smart Home devices.*** Participants also shared using their SPA to manage other smart home devices. In particular, six participants reported controlling other smart home devices. The devices included: smart bulbs, smart TVs and other smart speakers. In addition, some of our participants had tried to connect their SPA to other smart home devices they own but they did not succeed.

***Shopping.*** Four participants use the SPA for shopping. In particular, from those who use SPA for shopping, most of

them use SPA mainly to create shopping lists to later on purchase the items through the website or mobile application, as opposed to purchasing through the SPA. We explore the reasons for this in-depth in Section 5.4.

## 5.2 SPA Setup

From our 17 participants, 14 reported having set up their SPA while 3 stated that a partner or family member[2] had set up the device for use. All participants who set up the device stated that the setup process was easy and straightforward. They also stated that they used their personal Amazon and Google accounts to set up their devices. These were accounts that users were also using for other personal purposes such as shopping (Amazon accounts), and Android devices (Google/Gmail accounts). Both sets of users reasoned that it was easier and more convenient to use existing accounts than creating new ones, and that they preferred sharing it across the household rather than setting up multiple accounts — *"It is better to share one for convenience sake"* (P2). This is something that, to some extent, one could expect as it was already shown to happen in other home settings [33]. However, other participants reported that they wanted to link the devices with existing accounts and enjoy more of the added functionality and benefits SPA bring to them. As P10 put it *"so that its easy for me to see what's on my calendar"*, i.e., by linking the participant's existing account to their SPA they can set reminders that will sync with their regular calendar system. We found this particularly interesting, as it reinforces the importance of looking at the whole SPA ecosystem not just at the smart speaker placed in households.

### 5.2.1 Voice Recognition Setup

Although mostly used by the SPA for personalization rather than for security purposes, both Google Home/Assistant and Amazon Echo/Alexa offer voice recognition mechanisms for recognizing the voices of different users, so that they can tell users apart and personalize the interaction with them, named Voice Match [21] and Voice Profiles [2] respectively. In particular, Google users are given the opportunity to configure voice recognition as part of the initial setup process. Six Google users (6/7) setup voice recognition and reported that the device is usually able to distinguish their voices from the others, but with the mechanism being far from perfect, e.g. P12 said *"the times we have tested it seems that it can like but 70% of the time it doesn't seem perfect"*. In contrast, Amazon users are only given the chance to test the speech recognition process (ability to convert spoken words into a text and understand users' intent) as part of the initial setup, but not to configure voice recognition. Voice recognition (in this case Voice Profiles) can be set up at any point but always after the initial setup and as a separate process. Only 2 out of the 10 Amazon users reported completing voice recognition. Most users who did not set it up did not even know that this mechanism actually exists. Interestingly enough, some of those who did not complete voice recognition seemed not to understand or differentiate between speech recognition and voice recognition, and they would confuse them, thinking the SPA can distinguish between people without having set voice recognition. For those who understood the difference, they explained that speech recognition was a feature that allowed the device to recognize speech and change it to text while voice recognition involved the device being able to tell who is talking. When asked how the process works, they revealed that the device has an AI system which compared voices to distinguish between users. While these group of users reported that voice recognition is used to distinguish between users, some said it was for recognizing different accents (actually meaning speech recognition).

[2]Note here that we did not get into the tensions between those setting the devices and other household members, as this was already studied in-depth, including SPA too, in [46].

### 5.2.2 Third-party Skills Setup

Some third-party skills need to be setup either in terms of the permissions they need to access, e.g. smart speaker country and postcode for the case of Fuel Finder, or to link them to other online accounts to provide the functionality required, e.g., playing music through Spotify. We asked participants to describe the process of setting up the third-party skills they use. In some cases, this already started to shed light about their mental models. To setup the skill users share configuring their SPA to the skill they want to use. For example, P11 said: *"directly connected to my spotify account so it directly logs in to spotify and play music"*.

### 5.2.3 Connecting to Smart Home Devices

Managing other smart home devices through an SPA obviously requires connecting the SPA to the device. We asked participants to describe the process of connecting their SPA to their other smart home devices. They mentioned downloading the other smart home device mobile application and configuring it to their SPA, for example P3 stated *"I have the app on my phone so I use that to manage my activities between Alexa and the lamp"*. Other participants shared negative setup experiences, with some of them unable to connect their smart home devices to the SPA, with P9 stating *"I was unable to configure my smart TV, google home can't find the device"*.

### 5.2.4 Shopping Setup

Both Amazon Echo/Alexa and Google Home/Assistant support shopping lists by default, so users can just create lists and add items to buy. In terms of actually completing a purchase, Amazon allows users to optionally create a 4-digit pin code to be used when purchasing online. In particular, one of our

participants had set the code. Others who did not use their SPA for any shopping activities simply did not have the voice purchasing code setup and had it disabled. It should be noted that if a user has Amazon Voice Profiles (voice recognition) setup, they need to setup the pin code for purchasing but do not need to say it every time they want to complete a purchase.

## 5.3 Perceptions of SPA's Ecosystem

In order to explore what perceptions users had of the ecosystem, we considered all the four main use case scenarios and asked about different information-related activities (data processing, data storage, data sharing, data learning) and how they thought these activities where being conducted and where.

### 5.3.1 Data Processing

In general, our analysis shows that most SPA users believe that data collected by the device is processed locally in the device, though a few reported that the device needed to be connected to the Internet to work. Others explained that their requests are processed remotely and relayed back to the device. We explain this in detail below per type of use case, as there were some interesting differences worth mentioning across them.

***Built-in Skills.*** When asked to describe how the device processes and fulfills requests like weather updates, 10 out of the 13 participants that used built-in skills explained that the device locally processes these commands and respond to the user. For instance, one user described the device as a small brain, implying that the device listens and process commands before responding to the user. We also found a few participants who believe that the device communicates with the SPA provider to process commands and then responds to the user, but in many of these cases, this was because they thought the SPA connected with an online source of information to process requests. For instance, one participant mentioned that the device connected to the Google website for weather updates. P9 shared this *"... with the weather. I believe it comes from the Google site from their weather service"*.

***Third-party Skills.*** We observed that 10 out of the 11 participants who use third-party skills do not consider the third-party skills providers when describing how SPA process their request when a third-party skill is involved. While some users reported that data is sent to the SPA provider for processing, they did not mention any communication between the SPA provider and the third-party skill provider. This contrasts sharply with the very few participants who had a better understanding, though still incomplete and inaccurate, of how the process works when the SPA uses third-party skills. For instance, P8 stated *"well Alexa when I say I want to play a song she'll then connect to Spotify and search through the catalogue I guess then play the song"*.

***Managing Smart devices.*** We found that 5/6 participants believe that the smart speaker and other smart home devices communicate directly without involving other elements of the SPA architecture. For instance, when switching the smart lights on, they believe that the device communicates directly with the lights, implying that both the SPA provider and the smart light provider are not involved in any way. Some participants think these devices communicate through the mobile app (i.e., other smart home device's mobile app) installed in their mobile phones. For example, P2 said: *"basically Google Home talks to the light bulb via the mobile app installed, and they are connected via the network so I will say OK Google turn the light off/On and it will send the request to the application that controls the Philips light"*.

***Shopping.*** We found users shopping using the SPA talked about voice purchasing much in the same way as they would do for normal online purchases. For instance P5 said *"I just ask Alexa to add items to my basket"*. They also mentioned the SPA provider as being somehow involved in the process as the market/account they were buying from, e.g. P13 said *"once i ask alexa to add item to my basket she updates it on my Amazon account"*. While most of our participants did not complete the purchasing process using the device, we asked all users for their views concerning voice purchasing. The majority (13 out of the 17), reported having not thought about the process, but we observed that, similarly to those who use SPA to shop, their current online shopping practices influence their understanding of the voice purchasing process. They think about how they would select an item to buy, choose the method of payment and confirm the order.

### 5.3.2 Data Storage

In general, most users believe that their voice recordings, the history log and shopping history are all stored by the SPA provider. These users think this information is kept for building a profile about them, i.e. to understand their behaviour and interests. Regarding where data is stored, half of our interviewees believed that data collected by the SPA is stored locally in the smart speaker, while others reported that data is stored either in the cloud owned by the SPA provider, or both in the smart speaker and the cloud. One user stated that data is not stored at all since there is so much data to store. All our interviewees informed us that they do not know how data is stored and how long the provider keeps it. We further explain users' perceptions of how data is stored below depending on the use case.

***Built-in Skills.*** When using the device to complete everyday tasks like setting reminders, asking questions and requesting updates (e.g., weather), most participants believed that data is stored to learn more about them and personalize the SPA experience. Half of the participants believed that these data were stored locally in the device. One user stated such data was not stored at all since there were many data to store and the provider would not be able to handle it all. Another user mentioned that data (i.e., history) were stored in the mo-

bile app.

***Third-party Skills.*** Users who were using their SPAs with other third-party applications reported that their requests and history logs (e.g., playlists) were only stored by the SPA provider not mentioning the third-party provider. We observed that most users of third-party skills (9/11) do not mention their third-party skills providers storing any data.

***Managing Smart Devices.*** Most users who use their SPA to manage other smart home devices (5/6) reported that their commands directed to the smart home devices were stored, but they only assumed that it was to personalize and improve their SPAs. In terms of where or who stores the information, none of the participants mentioned any of the providers of the smart home devices they were managing through the SPA. They seem to only believe that their smart speaker stores all data locally in this case.

***Shopping.*** When using voice purchasing, users believe that their shopping lists and history are stored by the SPA provider. They reasoned that this is done to understand their shopping interests and behaviour. While the majority believed that this is for advertising purposes, some believed this is for improving the SPA. Regarding deletion of data, some participants stated that shopping history is immediately deleted from the device after shopping.

#### 5.3.3 Data sharing

While the usage of the SPA includes data being shared by the SPA provider and other different vendors or third-parties, we observed that users' perception of how data is handled and shared is mostly based on the stories of data misuse they know from other domains. For instance, users believe data is shared with data brokers and third-parties who are interested in influencing their behaviour, as P3 explained: *" ...so they would to try and influence users purchasing decisions"*. Other participants alluded to the Facebook and Cambridge Data Analytica case [45] and stated that they did not know with whom their data is being shared but believe it was being shared with other companies *P4 "they could give it to third-party people to target certain adverts to the user"*. However, some users reported that precisely because of recent data misuse incidents, they trusted their SPA providers not to share data with other parties.

In terms of the wider ecosystem, none of the participants who used the SPA with third-party skills (e.g., Uber) or with other smart devices (e.g., Phillips bulbs) mentioned data being accessible to these third parties (e.g. Uber or Phillips), let alone with whom these third parties might be sharing the data they gather. That is, no users mention the fact that, because they may have access to users' data because of how the SPA ecosystem work, that they could too share that data with others, not just the SPA provider. In terms of the specific data that participants believe is being shared, our participants informed us that their usage statistics, shopping habits and play-lists are being shared with other parties like advertisers.

#### 5.3.4 Data Learning

We also asked participants whether their SPA were capable of learning things about them based on their interaction or usage. In general, participants perceive SPA as intelligent and having the ability to learn new things about them without they telling them to the SPA. They describe them as a brain or having a memory to process and remember certain things about them. Others describe their device as an Artificial Intelligence (AI) system. In a similar way to processing and storing data, users seemed to attribute all the learning capabilities to the smart speaker as opposed to other parts of the SPA ecosystem involved in it, which they did not mention. They tended to personify the smart speaker and say it was intelligent.

Regarding how the device learns about them, 13/17 participants said the SPA analyses their usage patterns (i.e., questions, play-lists, history logs and shopping lists) to learn about their likes and dislikes. Our analysis shows that users believe their SPA are capable of learning about their shopping habits, their favourite music and radio stations, routines and its users. They also believe that the device uses what it learns about them to tailor adverts for them, serve them well, to influence their decisions and recommend better things to them (e.g., more music from their favourite artist), P17 *"It picks up adverts for example on my android phone I get adverts related to what I have asked my Google Home so it shows that element of the device listening"*. While P7 explained: *"I would probably imagine it stores your information and it [then] begins to predict through [the data] I would assume... some sort of like a pattern, therefore, it would [then] start to tailor things to people that fit that [particular] pattern."*

Some users have mixed attitudes toward the device being able to learn things about them; some perceive this as a negative trait while others see it as a useful feature, with some perceiving both depending on the context — reasoning very similarly to what well-known theories like Contextual Integrity [35] aim to explain. For instance, some users stated that the device being able to learn and know certain things about them (e.g., morning routine – favourite music, weather, traffic and news updates) is a good thing as it could simplify their life. However, they explained that it is not pleasant for the device to know sensitive things about them, for instance, health symptoms.

In general, users (including those that perceive learning as a good thing) find the idea that the device can learn about them being creepy, scary and invasive, sometimes because they could never tell when the device is doing the actual learning. P9 explained: *"In a way, it is good for it to give you suggestions. But, at the same time, it is scary because if you think about it, if it's learning things you are doing it is quite sinister. At the moment I am happy with it, but it does make me think about what information it can learn about me... what*

*profile it can build without me realising".*

## 5.4 Shopping

In our initial interviews, it quickly became apparent that one of the use cases we where considering, shopping, was actually worth studying more in-depth because of a seemingly low adoption by current SPA users. Both Amazon Echo and Google Home devices give users the opportunity to shop online. Therefore, we asked all participants about their shopping experience, the challenges and the concerns they have while using the device to shop. We aimed to understand how users view the process of shopping, from the moment they make the shopping list to the point of payment. We particularly sought to understand any differences in perception/use among those who do not use SPA to shop at all, those who use the SPA to aid their shopping even if not purchasing through the SPA (e.g. just using shopping lists), and those who actually use the SPA to purchase items.

Users view voice purchasing as a convenient way of shopping, with some tasks such as creating shopping lists and paying for goods faster than with other systems. We found that most of our participants (8/10 Echo users) had not set up voice purchasing code because they were not using voice purchasing features on their SPA. In general, most participants (7/10) told us that the voice purchasing code is a useful feature of the device and adds an extra layer of security. However, further analysis showed that most users are concerned that other people around the house (or neighbours) could hear the code and use it maliciously.

Below we summarize users' main struggles and concerns about using the SPA to shop. Mainly, we observed trust, or more specifically the *lack of trust*, emerging very strongly as a theme across different dimensions: products (visibility, comparisons, and mistakes), vendors, security of the connection, and privacy of the orders. In particular:

***Product visibility.*** When we asked our interviewees their thoughts on using the device for shopping, 10 out of 17 participants stated their biggest concern not being able to see the product they want to purchase.

For instance, P12 said: *"I [am] probably kind of against it, cause I will need [a] screen to see what I am buying, I need a lot of confirmation; how the products are and what I am buying, so a visual thing. So just using voice assistant I don't think I would ever do that.".*

***Product comparisons.*** Some users expressed the difficulties of comparing products when shopping using the device. Some Amazon Echo users stressed that Alexa did not give them a full description of the product but just the name and the amount. Other Amazon Echo users noted that one could not get the reviews of the product. Users also raised some concerns about fake products, that using the device one may end up ordering a fake product. For instance, P1 explained why he is not using voice purchasing: *"...erm only because I am aware of scams and fake products on Amazon, I would like to see what I am buying first."*

***Product mistakes.*** We found that some users were concerned about buying the wrong item because, sometimes, while they are creating the shopping list, the device gets the wrong item. In fact, two participants reported an unsuccessful attempt of shopping using the device as they mentioned the wrong items were added into their shopping basket. For instance, P5 said *"I just ask Alexa to add items to my basket and it does, but often it adds the wrong items..."*.

***Number and trustworthiness of vendors.*** Users expressed that they have a limited number of vendors to buy from than when shopping online. For instance, one Amazon user argued that they are limited when using the device because they cannot buy from other outlets. However, some Google Home users thought being connected to a single vendor (like Amazon Echo users) guarantees security as the user is just connected to a well-known and trusted outlet. Nonetheless, some Google Home users informed that the number of vendors is limited and there is a chance of not finding what they want. Other Google Home users stated that it is difficult to choose which vendor to use.

***Secure connection.*** Some users, mainly those who had not set up voice purchasing expressed their concerns over secure payment and connection during shopping. They stated that it was challenging to confirm whether they are connected to the right vendor or the payment process is secure. P9 said, *"... I don't know if the payment is secured or it's going through an encrypted site as a basic example, I like to see something on a screen rather than doing it on an automated home system."* Moreover, some further informed us that there are no visual cues to help them feel they are secure.

***People hearing orders and/or code.*** Some users who were not using the device to shop highlighted some privacy concerns of other people being able to hear what they are ordering, for instance, P14 stated *"people around you can easily hear your purchasing code which isn't safe if you think about it"*. Others said its easy for other members of the family or neighbours to hear what they are ordering and that can be unpleasant at times. They also mentioned concerns about others hearing the voice purchasing code and using it without their permission.

The above struggles and concerns make users utilize a number of strategies in order to minimize the concerns. Those include:

***Completing the order through the app.*** To avoid buying wrong items, some users stated that they use the device to create shopping lists but always confirm their orders before paying through mobile apps or website. For example, P5 said: *"I just ask Alexa to add items to my basket and it does ... and [then] I have to go to the app to make the purchase."* Most users who used this strategy mentioned that the device is good for making shopping lists but not ideal for shopping especially when product details matter.

***Disabling voice purchasing.*** Most participants mentioned that they decided not to enable voice purchasing because they do not trust it. These were users who earlier revealed that they were not sure how secure the device is when shopping.

***Shopping through other platforms.*** Some users explained that they still prefer to shop using their apps or the web. They explained that shopping using other platforms gave them the opportunity to find better deals and get the right items. Some users further explained that these other platforms are trustworthy and have been using them for some time. For example, P10 noted: *"I would say the device is great for other things, but in terms of shopping, it is useful to add things to your basket, but I would say its better to buy through the website or app, so you know its safe and secure."*.

## 5.5 Threat Models and Coping Strategies

To understand users' privacy and security concerns regarding owning and using SPAs, we asked users if they thought their devices could be exploited maliciously or if their data could be at risk while using the device. We were also interested in the threat agents – actors who might be interested in such attempts. Considering the size of the ecosystem and the number of stakeholders involved, these questions aimed at getting participants to describe the threats and the attacks that SPAs might be subjected to. After these questions, we wanted to know what users do to protect themselves from these threats and attacks.

All of our participants reported that their SPA could be exploited. They described how different threat agents could attack the device. In general, we observed many gaps in their threat models; users consider few threat agents and exclude the people they share the device with. Also, they do not consider malicious skills or SPA providers. Users are mostly worried about unwanted listening from the device. They reported not knowing how to protect themselves or their devices from technical attacks but shared various non-technical solutions they develop to protect themselves.

### 5.5.1 Threat Agents

While some of the users explained that anyone could *hack* the SPA, the most common threat agents that users discussed were: hackers, government agencies and data brokers (advertisers). Many of our participants used words like "criminals" and "fraudsters" to describe potential threat agents. We grouped all these under the theme "Hackers". Users gave various reasons, i.e. *motivations*, to why these threat agents would be interested in attacking the SPA. They mentioned that hackers (and fraudsters) would be interested in targeting SPA for financial gain, to get personal data which they can then sell and for blackmailing purposes; government might do it for spying on users and influencing their decisions; and advertisers would do it for understanding users' usage behavior and use that for marketing purposes. We also found that participants who mentioned advertisers highlighted that data generated by users is considered important, and everyone is interested in it. However, most users who mentioned "fraudsters" and "criminals" linked them to financial gains. For instance, P4 stated *"... with the shopping feature [available], potential people who want to steal money of you can target it... because your credit card is stored so I would say fraudsters"*.

Despite recent news, e.g., Amazon releasing a user's Amazon echo recordings to another user [44], users do not normally consider the SPA providers or providers who have access to the data, e.g. third-party skill providers, as threat agents. Also, no one mentioned other household members as a potential source of problems. However, studies suggest that smart home devices are weaponized within the family [23]. The only hint towards this was a few participants who mentioned the problem of other household members and neighboors overhearing the voice purchasing code.

### 5.5.2 Attack Types

We found that while users' threat models consist of different threat agents, many users struggled to describe attacks that SPAs can suffer. The most prevalent attack mentioned by our users is unwanted listening. All our users raised this as a concern and mentioned different threat agents hacking the device to listen and spy on them. Some users shared advanced attacks such as attacking the network the device is connected to and hijacking the commands, but still attacks did not normally correspond to the real attack surface of SPA (see Section 3). For example P17 shared *"They are connected to the network so they have IP and storage so they can become part of a botnet"*. Also, they hardly related to any parts of the SPA ecosystem but the smart speaker — e.g. participants did not consider malicious skills [29].

### 5.5.3 Coping Strategies

We found that users do not take any technical solutions to deter threats or protect themselves. We now discuss the strategies they follow to protect themselves below.

***Unable to protect themselves***. Many participants reported that it is difficult to protect the device because they do not know what attacks might affect their devices. P4 said: *"With these sorts of things, I don't really know if there is a way of protecting yourself..."*. P1 further explained: *"With my PC and phone, I have an anti-virus [installed], but I don't know how you could protect a speaker..."*. This is remarkable, as it shows many users, even if they might do something to better protect themselves, simply cannot do it because they do not know what to do.

***Not enabling certain features.*** Users reported that they disable (or do not set up) some features and functionality of

the SPA to minimize or avoid being at risk. One example of this, as mentioned above, is disabling voice purchasing to avoid risks associated with shopping. This means many users are just restricting themselves in terms of the SPA capabilities they could be using. P10 said: *"Somehow yes, I would limit the things I use it for like I wouldn't use it for purchasing at all I'll stick to shopping lists."*

***Using other devices.*** Some users reported that they use other devices to complete specific tasks in order to minimize what the device knows about them. For instance, P9 said: *"...checking the weather would be ok, but I would be concerned, for example, if I wanted to find out about a certain illness a family member has, I wouldn't do it through Google home...I would use the computer [be]cause I don't want that to be stored [in the SPA]"*. Another example of this is that, as mentioned above, participants tend to complete purchases using other devices like mobile apps after having created a shopping list with the SPA. Again, this means that users are simply not using the SPA for tasks it is capable of doing.

***Turning off or muting the smart speaker.*** We found that some users switch off their SPAs to stop them from listening. They turn the device off when they are sleeping, having private conversations and when they are not home to avoid unauthorized people using them. P9 explained: *"...I would turn it off when we are not in the house so people can't access it when we are not in."* This finding confirms what was also found in [30], where they asked about whether users used the muting button of smart speakers, which in turn revealed that many users were turning off the speakers altogether.

## 6 Discussion

We now discuss our findings, their implications, and some recommendations.

### 6.1 More than Smart Speakers

The majority of users see the smart speakers as the place from the whole SPA ecosystem where most of the data processing, storage and learning happens. For instance, when asked to describe how the SPA process and fulfill requests like weather updates, the majority of our participants explained that the device locally processes these commands and respond to the user, mainly limiting the SPA to the smart speaker, which would in turn be some kind of a small brain. This shows that most users have a very simple and inaccurate mental model of the SPA ecosystem. Even those who actually recognize that the SPA needs to search for and find information online do have incomplete mental models. Very few participants clearly involved the SPA provider in the processing, storage, sharing and learning capabilities of SPA, let alone other important actors in the ecosystem like the third-party skills providers and the vendors of smart home devices they manage through the SPA. Therefore, better awareness and transparency mechanisms may help users understand how SPA operate, not necessarily from a technical point of view, but just enough to understand the implications in terms of their data. Awareness and transparency mechanisms, however, need to be engineered carefully, to avoid these mechanisms becoming a lot of information to digest, which may intimidate and/or become a burden on the users, ultimately ending up of not much use. In fact, some participants actually complained about SPA privacy policies and terms of service not being clear enough for them to understand how their data is handled, something that one would expect as it is the case in other domains [13, 25, 38, 40]. Recent research suggested that privacy notices should be relevant, actionable and understandable [41]. In particular, the authors identify four main dimensions to consider when designing to provide notice: timing, when should a notice be presented; channel, how should the notice be delivered; modality, how the information should be conveyed; and control, how choice options are integrated into the notice. Research on SPA notices exploring those dimensions would be really interesting, particularly as conveying notice across the SPA ecosystem, considering its complexity and the actors involved, is non-trivial.

### 6.2 What do I do to protect myself?

Having better transparency mechanisms that help improve users' mental models of the SPA ecosystem may not necessarily mean that users are able to protect themselves better. Although most participants were clearly unaware of the potential threats, which could mean that they underestimate the security and privacy risks of SPA, and one might be tempted to attribute this to the inaccurate and incomplete mental models users have, one of the main problems we encountered is that most users simply did not know what they could do to protect themselves when using SPA. This actually leads to a situation whereby users minimize the use they make of the SPA to just the cases they think (whether actually right or wrong) are less dangerous. If users are to make the most of SPA, we definitely need more usable security and privacy mechanisms that seemingly integrate with the SPA ecosystem, together with the awareness and transparency mechanisms already mentioned, which in turn may help increase users' trust on SPA.

**AI to personalize security/privacy.** The first example of potential mechanisms to explore would be those that could leverage the AI capabilities SPA have to personalize the experience to users, so that they would be used to personalize users' security and privacy experience. This would contribute to the cases we found participants felt SPA learning is a good thing. In this way, recent research already suggested variables to consider for permissions within a household [24]. This, together with permissions across the entire SPA ecosystem considering the actors involved, could be the basis for SPA to learn what are the kind of contextual social norms that apply

for particular users and households to govern data processing, storage, sharing, and learning based on the context to help users manage and control their data across the SPA ecosystem. In fact, the feasibility of learning contextual social norms was already shown in other domains [7, 11, 18, 42], and more recently in generic smart homes [4], but this still needs to be considered in the context of the whole SPA ecosystem.

**Voice recognition for usability and as building-block.** We observed that when the voice recognition setup process is included in the initial SPA setup as with Google Home/Assistant, many more participants seemed to configure it and, actually, they found the process easy and straightforward. In contrast, voice recognition in Amazon Echo/Alexa is not part of the initial setup, and the vast majority of users had not tried to set voice recognition, with some of them not even knowing the mechanism exists. While voice recognition may still not be a mature-enough authentication mechanism in SPA, as it has been shown to be vulnerable to attacks such as spoofing through replay attacks [9], there is indeed ongoing research to make it more secure [16]. The good news is that, from a usability point of view, this looks like an interesting research line, because of the aforementioned proportion of participants who went for voice recognition when they knew about it and were given the chance to set this up at the initial setup stage. Voice recognition could also be the basis for other security mechanisms or to increase trust in some SPA use cases such as shopping, as explained next.

## 6.3 Trusted Shopping

We found a lack of trust from users when shopping using the SPA. While some participants found it useful and convenient to use some of the SPA's shopping capabilities such as shopping lists, participants would not normally purchase the items through SPA. We identified that the main cause of this was that users did not trust the products, the vendors, and the process, including the security of the connections and whether other people might be able to overhear their purchases and purchase codes. These trust issues need to be addressed in order to foster purchases through SPA, even more if we look towards a future where we will delegate more and more tasks to SPA [10]. Research on the particular mechanisms to make purchasing through SPA more trustworthy seems like an exciting line of future research. For instance, in terms of products and vendors, novel ways for an SPA to somehow provide more verbal information about the products and the vendors, such as product reviews or vendors' reputation, would need to be engineered in a usable way. Also, this type of assurances might need to be complemented with other modalities, something that may be easier with the new generation of multi-modal smart speakers, such as the new 2nd generation of Amazon echo, which includes a screen [3] users could use to check the products in the shopping list to purchase all in one place, with the SPA quickly ordering the items as soon as the user confirms verbally. Also, and as mentioned above, having voice recognition from the beginning would make it so the voice purchasing codes needed in Amazon would not need to be repeated in each purchase (as it is actually the case [2]), also mitigating the concerns some users had in terms of others overhearing the code and using it.

## 6.4 Limitations

The methodology used was mostly qualitative and exploratory in nature, therefore the hypotheses we formulated based on our findings, emerging themes and discussion coming from the grounded-theoretic analysis, would obviously need to be tested in a follow-up confirmatory study to assess their validity and generalizability. We focused on current SPA users, so we could explore the ecosystem and the parts they understand or use more/less and why, e.g. lack of trust regarding SPA shopping, and because previous work [30] had already looked at users and non-users of SPA to study reasons for adoption. Finally, the interviews were conducted Jan-Feb 2019 after some major news, including Amazon sending thousands of recordings to the wrong user [44], but before the most recent news regarding Alexa recordings being analysed by humans. While this might alter mental models regarding the SPA provider, sometimes mentioned by participants, it might not regarding third-party skills or other third-parties, who were not in the news and hardly mentioned by participants. Nevertheless, understanding how such news could alter users' mental models, particularly in terms of the SPA provider, should be studied.

# 7 Conclusion

This paper reports our study of users' perceptions of the SPA ecosystem through semi-structured interviews around four main use cases of SPA (built-in skills, third-party skills, managing other smart home devices, and shopping). We uncovered users' misunderstanding of SPA ecosystem, with most users showing a very limited conception of SPA and inaccurate and incomplete mental models of the SPA ecosystem and related data activities (processing, storing, sharing, and learning). We also uncovered the lack of trust users have with some of the use cases of SPA, and how this is hampering adoption particularly of purchasing through the SPA, with users not having enough information to assess their trust in the products, the vendors, and the process of voice purchases. In addition, we reported the threat models users have of SPA, showing both threat agents and types of attacks users consider possible. We also show the mainly non-technical coping strategies users follow to try to protect themselves. Finally, we presented design implications for how SPA might support users' expectations and needs with regards to privacy and security, including researching on mechanisms that help increase awareness, transparency, control, and trust across the SPA ecosystem.

## References


[1] The alexa skill store for france is a fast growing land of opportunity, https://voicebot.ai/2018/11/03/the-alexa-skill-store-for-france-is-a-fast-growing-land-of-opportunity/, 2018.

[2] Amazon. *About Alexa Voice Profiles*, 2019 (accessed February 22, 2019). https://www.amazon.com/gp/help/customer/display.html?nodeId=202199440.

[3] Amazon. *All-new Echo Show (2nd Gen) – Premium sound and a vibrant 10.1" HD screen*, 2019 (accessed February 22, 2019). https://www.amazon.com/All-new-Echo-Show-2nd-Gen/dp/B077SXWSRP.

[4] N. Apthorpe, Y. Shvartzshnaider, A. Mathur, D. Reisman, and N. Feamster. Discovering smart home internet of things privacy norms using contextual integrity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(2):59, 2018.

[5] S. Bay. Ai assistants are poised for major growth in 2018. 2018.

[6] Bryman. *Social research methods*. Oxford university press, 2015.

[7] G. Calikli, M. Law, A. K. Bandara, A. Russo, L. Dickens, B. A. Price, A. Stuart, M. Levine, and B. Nuseibeh. Privacy dynamics: Learning privacy norms for social software. In *Proceedings of the 11th International Symposium on Software Engineering for Adaptive and Self-Managing Systems*, pages 47–56. ACM, 2016.

[8] K. Charmaz. *Constructing grounded theory*. Sage, 2014.

[9] S. Chen, K. Ren, S. Piao, C. Wang, Q. Wang, J. Weng, L. Su, and A. Mohaisen. You can hear but you cannot steal:defending against voice impersonation attacks on smartphones. *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*, 2017.

[10] P. Cohen, A. Cheyer, E. Horvitz, R. El Kaliouby, and S. Whittaker. On the future of personal assistants. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 1032–1037. ACM, 2016.

[11] N. Criado and J. M. Such. Implicit contextual integrity in online social networks. *Information Sciences*, 325:48–69, 2015.

[12] T. Denning, T. Kohno, and H. M. Levy. Computer security and the modern home. *Communications of the ACM*, 56(1):94–103, 2013.

[13] J. B. Earp, A. I. Anton, L. Aiman-Smith, and W. H. Stufflebeam. Examining internet privacy policies within the context of user privacy values. *IEEE Transactions on Engineering Management*, 52(2):227–237, May 2005.

[14] J. S. Edu, J. M. Such, and G. Suarez-Tangil. Smart Home Personal Assistants: A Security and Privacy Review. *arXiv eprint arXiv:1903.05593*, 2019.

[15] P. Emami-Naeini, H. Dixon, Y. Agarwal, and L. F. Cranor. Exploring how privacy and security factor into iot device purchase behavior. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, page 534. ACM, 2019.

[16] H. Feng, K. Fawaz, and K. G. Shin. Continuous authentication for voice assistants. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, pages 343–355. ACM, 2017.

[17] E. Fernandes, J. Jung, and A. Prakash. Security analysis of emerging smart home applications. In *Security and Privacy (SP), 2016 IEEE Symposium on*, pages 636–654. IEEE, 2016.

[18] R. Fogues, P. K. Murukannaiah, J. M. Such, and M. P. Singh. Sharing policies in multiuser privacy scenarios: Incorporating context, preferences, and arguments in decision making. *ACM TOCHI*, 24(1):5, 2017.

[19] N. Fruchter and I. Liccardi. Consumer attitudes towards privacy and security in home assistants. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, page LBW050. ACM, 2018.

[20] B. G. Glaser and A. L. Strauss. *The discovery of grounded theory: Strategies for qualitative research*. Transaction publishers, 2009.

[21] Google. Set up multiple users for your speaker or smart display. https://support.google.com/assistant/answer/9071681, 2017. Last accessed 20-February-2018.

[22] W. Haack, M. Severance, M. Wallace, and J. Wohlwend. Security analysis of the amazon echo. *Allen Institute for Artificial Intelligence*, 2017.

[23] M. Hansen and B. Hauge. Scripting, control, and privacy in domestic smart grid technologies: Insights from a danish pilot study. *Energy research & social science*, 25:112–123, 2017.

[24] W. He, M. Golla, R. Padhi, J. Ofek, M. Dürmuth, E. Fernandes, and B. Ur. Rethinking access control and authentication for the home internet of things (iot). In *27th USENIX Security Symposium (USENIX Security 18). USENIX Association, Baltimore, MD*, 2018.

[25] P. G. Kelley, J. Bresee, L. F. Cranor, and R. W. Reeder. A nutrition label for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security*, page 4. ACM, 2009.

[26] V. Kepuska and G. Bohouta. Next-generation of virtual personal assistants (microsoft cortana, apple siri, amazon alexa and google home). In *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, pages 99–103. IEEE, 2018.

[27] B. Kinsella. The Information Says Alexa Struggles with Voice Commerce But Has 50 Million Devices Sold. https://voicebot.ai/2018/08/06/the-information-says-alexa-struggles-with-\voice-commerce-but-pass, 2018. Last accessed 28-February-2019.

[28] D. Kumar, R. Paccagnella, P. Murley, E. Hennenfent, J. Mason, A. Bates, and M. Bailey. Skill squatting attacks on amazon alexa. In *27th {USENIX} Security Symposium ({USENIX} Security 18)*, pages 33–47. USENIX, 2018.

[29] D. Kumar, R. Paccagnella, P. Murley, E. Hennenfent, J. Mason, A. Bates, and M. Bailey. Skill squatting attacks on amazon alexa. In *27th USENIX Security Symposium (USENIX Security 18)*, pages 33–47, Baltimore, MD, 2018. USENIX Association.

[30] J. Lau, B. Zimmerman, and F. Schaub. Alexa, are you listening?: Privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):102, 2018.

[31] X. Lei, G.-H. Tu, A. X. Liu, C.-Y. Li, and T. Xie. The insecurity of home digital voice assistants-amazon alexa as a case study. *arXiv preprint arXiv:1712.03327*, 2017.

[32] E. Luger and A. Sellen. "like having a really bad pa": The gulf between user expectation and experience of conversational agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 5286–5297, New York, NY, USA, 2016. ACM.

[33] T. Matthews, K. Liao, A. Turner, M. Berkovich, R. Reeder, and S. Consolvo. She'll just grab any device that's closer: A study of everyday device & account sharing in households. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 5921–5932. ACM, 2016.

[34] A. Mutcher. Google assistant app total reaches nearly 2400 thats not real number really 1719, 2019. Last accessed 22-Feb-19.

[35] H. Nissenbaum. Privacy as contextual integrity. *Wash. L. Rev.*, 79:119, 2004.

[36] N. Nthala and I. Flechais. Informal support networks: an investigation into home data security practices. In *Fourteenth Symposium on Usable Privacy and Security ({SOUPS} 2018)*, pages 63–82, 2018.

[37] OVUM. Virtual digital assistants to overtake world population by 2021. 2017.

[38] I. Pollach. What's wrong with online privacy policies? *Communications of the ACM*, 50(9):103–108, 2007.

[39] S. T. S. Portal. *Worldwide intelligent/digital assistant market share in 2017 and 2020, by product*, 2019 (accessed Feb 22, 2019). https://www.statista.com/statistics/789633/worldwide-digital-assistant-market-share/.

[40] K. M. Ramokapane, A. C. Mazeli, and A. Rashid. Skip, skip, skip, accept!!!: A study on the usability of smartphone manufacturer provided default features and user privacy. *Proceedings on Privacy Enhancing Technologies*, 2019(2):209–227, 2019.

[41] F. Schaub, R. Balebako, and L. F. Cranor. Designing effective privacy notices and controls. *IEEE Internet Computing*, 21(3):70–77, May 2017.

[42] Y. Shvartzshnaider, S. Tong, T. Wies, P. Kift, H. Nissenbaum, L. Subramanian, and P. Mittal. Learning privacy expectations by crowdsourcing contextual informational norms. In *Fourth AAAI Conference on Human Computation and Crowdsourcing*, 2016.

[43] M. Singleton. Alexa can now set reminders for you. https://www.theverge.com/circuitbreaker/2017/6/1/15724474/alexa-echo-amazon-reminders-named-timers, 2017. Last accessed 28-February-2019.

[44] N. Statt. *Amazon sent 1,700 Alexa voice recordings to the wrong user following data request*, 2018 (accessed Jan 22, 2019). https://www.theverge.com/2018/12/20/18150531/amazon-alexa-voice-recordings-wrong-user-\gdpr-privacy-ai.

[45] A. Valdez. *Everything You Need to Know About Facebook and Cambridge Analytica*, 2018 (accessed Jan 22, 2019). https://www.wired.com/story/wired-facebook-cambridge-analytica-coverage/.

[46] E. Zeng, S. Mare, and F. Roesner. End user security and privacy concerns with smart homes. In *Thirteenth Symposium on Usable Privacy and Security ({SOUPS} 2017)*, pages 65–80, 2017.

[47] S. Zheng, M. Chetty, and N. Feamster. User perceptions of privacy in smart homes. *arXiv preprint arXiv:1802.08182*, 2018.

## A Screening Questions

1. Which device do you own?
   Amazon Echo
   Google Home
   Apple Homepod
   Microsoft Cortana
   2. How long have you been using the device?
   3. How many people within your household use the device?
   4. Which of the following voice commands are used to awaken Amazon's personal assistant?
   "Alexa"
   "Computer"
   "Hey Amazon"
   "I don't own this device"
   5. Which of the following voice commands are used to awaken Google personal assistants?
   "Hey Google"
   "Ok Google"
   "Google"
   "I don't own this device"
   6. Do you use any of the following services on your device?
   Play music
   Set alarm and reminders
   Shopping
   Third party services
   Managing other smart home devices
   7. "Amazon Echo supports third party services called skills"?
   True
   False
   I don't own this device
   8. "Google Home supports third party apps"
   True
   False
   I don't own this device
   9. Which device has voice purchasing code?
   Amazon Echo
   Google Home
   I don't know
   10. Which device has the capability to distinguish between different speakers?
   Amazon Echo
   Google Home
   I don't know

## B Interview Questions

1. Which device do you own?

2. Can you tell me about your device, what made you start using it?
   Follow-up: How long have you been using it?
   Follow -up: Other than you, who else uses it?
   3. What do you use the device for?
   Follow up: Do you use third party skills/apps?
   Follow up: What do you use?
   Follow up: How often do you use it?
   Follow up: Did you have to setup anything before you started using it?
   4. Other than your device, do you own any other smart home device?
   Follow up: Do you use your device to control your other smart home device?
   Follow up: How useful is your device in terms of controlling your smart home device?
   5. How did you register your device?
   Follow up: Was this done with your existing account? [If used an existing one]
   Follow up: Is this just for your device or you use the account for other things as well?
   Follow up: Can you tell me why you linked them?
   [If not linked]
   Follow up: Is there any reason why you didn't link them?
   [If created new account]
   Follow up: Is there a reason behind creating a new account than using an existing one?
   6. How many accounts do you have setup on your device?
   Follow up: Do these belong to others that use the device?
   Follow up: Do you use those other accounts or just one?
   [If only one account]
   Follow up: Is this shared by multiple users?
   Follow up: Can you explain why you chose to share an account?
   7. Have you completed the voice recognition process?
   Follow up: How did you find it?
   Follow up: Does the device respond to you when you speak to it?
   Follow up: When the device doesn't respond or understand you, what do you do?
   8. Can the device distinguish users or tell users apart?
   Follow up: How do you think this process works?
   Follow up: Did you experience any challenges in terms of the device identifying who you are?
   Follow up: If any, what did you do to overcome it?
   Follow up: Did you do anything to make the device recognize and identify your voice?
   Follow up: What did you do?
   9. Do you use the device to shop?
   Follow up: Can you share with me your experience in using the device to shop?
   Follow up: What do you exactly do when you shop using the device?

Amazon Echo Users only: Did you setup the voice purchasing code?

Follow up: Can you describe your experience setting up your purchasing code?

Follow up: Is your voice purchasing code always enabled?

[If disabled]

Follow up: Can you tell me why you have it disabled?

Follow up: What are your thoughts on shopping using the device?

10. Scenario 1 - Built in Services

When[NAME OF BUILT IN SERVICES] how does the device get the information you requested?

Follow up: Do you know if these requests are stored?

Follow up: [If yes] where do you think they are stored and for how long?

Follow up: Do you think Amazon or Google use this data for any purposes?

Follow up: Do you think Amazon or Google share your data with third parties like advertisers?

Scenario 2: Managing other smart home devices

You have mentioned that you use your device to manage your other smart home device, can you describe to me how you think this process works?

Follow up: Do you think what you do [activity history] are stored?

Follow up: [If yes] where do you think they are stored and for how long?

Follow up: Do you think your device shares data with [NAME OF THE OTHER DEVICE COMPANY]?

Follow up: Do you think Amazon or Google and [NAME OF THE OTHER SMART HOME DEVICE] share your data with other third parties such as advertisers?

Scenario 3: Third Party Apps You have mentioned that you use third party skills/apps on your device [NAME] can you describe to me how you think this process works?

Follow up: How does Alexa or [Google] communicate with [NAME OF App]?

Follow up: Do you know if these requests are stored and for how long?

Follow up: Do you think the device shares date with [NAME OF THE SKILL/APP]?

Follow up: Do you think Amazon or [Google] and [NAME OF THIRD PARTY SKILL/APP] share your data with other third party companies such as advertisers?

Scenario 4: Voice Purchasing

You mentioned that you sometimes use your device to purchase online, can you briefly describe to me how you think this process works?

Follow up: Do you know if purchasing orders are stored?

Follow up: [if yes] where do you think they are stored and for how long?

Follow up: Do you think Amazon or Google use this data for any purposes?

Follow up: Do you think this data is shared with other third parties like advertisers?

11. DO you think the device is able to learn things about you based on what you have asked before?

Follow up: How do you think the device is able to do that?

Follow up: Can you give me an example of what you think it has learned about you previously

Follow up: What are your thoughts on the device being able to learn things about you that you may not have said to it explicitly?

Follow up: Where do you think what the device learns about you is stored?

Follow up: Do you think Amazon or Google could use what the device learns about you for any purposes?

Follow up: Do you think what the device learns about you is shared with third parties e.g. advertisers?

12. Do you think the device could be exploited maliciously?

[If yes]

Follow up: Who do you think would be interested in exploiting the device?

Follow up: What do you think their motive is?

[If no]

Is there any specific reason you think it cannot be exploited?

13. How do you protect yourself from the device or those who might attack it?

Follow up: How effective is that?

14. Do you have any concerns on how the device handles your data?

[If any concerns]

Follow up: Does that impact the way you use the device?

15. Have you ever experienced any conflicts with others that have access to your device?

16. Have you previously experienced any incidents were the device has done something without you activating it?

17. Is there anything else you do apart from what we have talked about already?