Expecting the Unexpected: Understanding Mismatched Privacy Expectations Online

Ashwini Rao, Florian Schaub, Norman Sadeh, and Alessandro Acquisti,
Carnegie Mellon University; **Ruogu Kang,** ***Facebook***

https://www.usenix.org/conference/soups2016/technical-sessions/presentation/rao

This paper is included in the Proceedings of the
Twelfth Symposium on Usable Privacy and Security (SOUPS 2016).

June 22–24, 2016 • Denver, CO, USA

ISBN 978-1-931971-31-7

Open access to the Proceedings of the
Twelfth Symposium on Usable Privacy
and Security (SOUPS 2016)
is sponsored by USENIX.

Expecting the Unexpected: Understanding Mismatched Privacy Expectations Online

Ashwini Rao
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA, USA
arao@cmu.edu

Florian Schaub
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA, USA
fschaub@cs.cmu.edu

Norman Sadeh
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA, USA
sadeh@cs.cmu.edu

Alessandro Acquisti
Heinz College
Carnegie Mellon University
Pittsburgh, PA, USA
acquisti@andrew.cmu.edu

Ruogu Kang
Facebook
Menlo Park, CA, USA
ruoguk@fb.com

ABSTRACT

Online privacy policies are the primary mechanism for informing users about data practices of online services. In practice, users ignore privacy policies as policies are long and complex to read. Since users do not read privacy policies, their expectations regarding data practices of online services may not match a service's actual data practices. Mismatches may result in users exposing themselves to unanticipated privacy risks such as unknowingly sharing personal information with online services. One approach for mitigating privacy risks is to provide simplified privacy notices, in addition to privacy policies, that highlight unexpected data practices. However, identifying mismatches between user expectations and services' practices is challenging. We propose and validate a practical approach for studying Web users' privacy expectations and identifying mismatches with practices stated in privacy policies. We conducted a user study with 240 participants and 16 websites, and identified mismatches in collection, sharing and deletion data practices. We discuss the implications of our results for the design of usable privacy notices, service providers, as well as public policy.

1. INTRODUCTION

Privacy policies serve as the primary mechanism for notifying users about a website's data practices, such as collection and sharing of personal information. However, website privacy policies, written in natural language, can be long, time consuming to read [18, 30], and difficult to understand for users [42, 46]. They are therefore often ignored by users [9, 43]. One approach for helping users is to provide additional privacy notices that are based on privacy policies, but are shorter, easier to understand and more usable [10,22,49,55]. Prior work on privacy notices has focused on summary notices that display data practices in an easy to understand visual format [10, 22, 49, 55]. Even with simplified privacy notices, much of the information may not be relevant to users. Many data practices are expected and obvious, may not create concern, or do not apply to the user's current interaction with a service. For instance, it is likely obvious to users that when they explicitly provide their contact and payment details to an online store that that information will be collected and used to fulfill the purchase. However, data practices that are unexpected may result in a loss of trust and a sense that one's privacy has been violated, even if the practices in question were disclosed in the service's privacy policy [47]. More importantly, expectations influence decision making [17] and mismatches between users' expectations and website data practices may lead to incorrect privacy-related decisions.

The framework of contextual integrity highlights the impact of social context and information type on flow of information [34,35]. Expectations regarding flow of information may vary by social context and information type. For instance, collection of financial information on a banking website may be more expected than collection of health information. Privacy expectations are further influenced by an individual's personal, social and cultural background, as well as expectations in social roles and other "borders" that delineate spheres of privacy [29, 39]. For instance, depending on their technical knowledge, some users may expect that websites they visit can infer their rough location based on their IP address. For others, inference of their location may be completely unexpected.

Although unexpected data practices may be described in a privacy policy, they are likely to be overlooked among descriptions of practices that are expected or irrelevant to the user's current transactional context. The verbosity of privacy policies may be necessary to comply with legal and regulatory requirements, but it also means that privacy policies are not helpful to users in making informed privacy decisions [9]. In order to provide transparency to users, compliance-oriented privacy policies should be complemented with short form notices tailored to the user's transactional context [49] that should warn users about unexpected prac-

Symposium on Usable Privacy and Security (SOUPS) 2016, June 22–24, 2016, Denver, Colorado.

tices in particular [14]. The challenge, however, lies in identifying unexpected practices. Users' privacy preferences have been studied in different contexts [23, 38, 51]. However, privacy behavior differs from stated preferences [36], and preferences are not reliable for identifying mismatches between privacy expectations and a company's actual practices.

1.1 Contributions

To advance toward more practical solutions that can impact privacy notice design, we propose a practical approach for determining mismatches between users' expectations and services' data practices, as stated in their privacy policies. Research in other fields, such as marketing, has highlighted that the term "expectations" can mean at least four different things in consumers' context [32], but in the privacy context most work has focused on expectations in the desired sense or preferences [23,33], or has not clarified the meaning of expectation [13, 16, 26]. We propose to elicit privacy expectations, in the sense of "expected occurrence likelihood," rather than aspirational privacy preferences, and use the elicited expectations to identify mismatches with stated data practices. By focusing on expectations of what is happening, we avoid problems with unreliable subjective preferences of what should happen.

We compared expectations elicited from users with website data practices extracted from website privacy policies with manual annotations. Our analysis shows that characteristics of a website, such as its type, as well as user characteristics, such as privacy knowledge and concern, are strong predictors of data practices that are likely to be unexpected.

From our results, we derive guidelines on what data practices are likely unexpected and should therefore be emphasized in privacy notices. Knowledge about which characteristics affect expectations can be used to contextualize notices to the type of website and transactional context, as well as personalize notices to specific audiences in order to make unexpected practices more salient compared to expected practices, and thus make it easier for users to obtain information relevant for making informed privacy decisions. Our insights can benefit third-parties that generate simplified privacy notices, for example via browser extensions, as well as service providers. Both can use our approach to identify data practices that users will likely not expect and may cause privacy concern. Service providers could assuage user concerns by explaining the rationale behind such data practices.

While we manually extracted data practices from privacy policies to ensure reliable ground truth data, recent advances in the semi-automated analysis of privacy policies [6, 25, 48, 53, 55] show promise that our approach can be automated and scaled up to a large number of websites once such techniques are sufficiently robust.

2. BACKGROUND & RELATED WORK

In the United States, website privacy policies serve as the dominant mechanism for informing Internet users about website data practices such as collection, sharing and retention of personal information [47]. A website privacy policy, written in natural language, contains statements about the website's data practices. Policies may be long and time consuming to read [30] and require high language proficiency skills [18], which can lead to differences in how the general public and legal scholars interpret policy statements [46].

One approach to help users understand website data practices is to provide more concise privacy notices in addition to privacy policies [49]. Such privacy notices may be based on privacy policies, but are generally shorter and more usable. They could be provided either by website operators or by third parties. Research on privacy notices has focused on display formats that are easier for users to understand [10, 22, 31, 48, 55].

Users' privacy preferences and willingness to share information have been studied in many contexts [2, 23, 38, 51]. Acquisti et al. [2] note that privacy preferences and privacy decision making are prone to uncertainty, context-dependent, shaped by heuristics and cognitive biases, malleable and easily influenced by framing. Elicited privacy preferences can therefore be difficult to generalize, and actual behavior often deviates from stated preferences [36]. Observing privacy behavior is preferable, but behavioral studies can be challenging and resource-intensive to conduct at scale.

Privacy research has also explored the concept of expectations of privacy, including seminal work by Altman [3, 28], Marx [29] and Nissenbaum [34, 35]. For instance, Altman showed that individuals continuously modify their behavior to achieve an expected level of privacy [3]. Nissenbaum discusses how expectations of privacy are shaped by context [34]. However, to the best of our knowledge, privacy research has not focused on the potential for multiple levels or types of expectations. For example, in Altman and Nissenbaum's work, there is a single notion of expectation that may change based on different factors such as context. Privacy research typically differentiates between expected privacy and actual privacy, for example, Altman differentiates between desired and achieved levels of privacy [3].

However, research in other domains indicates that individuals have multiple levels or types of expectations [15,32,50,54] and these types of expectations can impact constructs such as consumer satisfaction [50] and performance [15]. For instance, Miller distinguishes four expectation types: *Ideal, Expected, Minimum Tolerable,* and *Deserved* [32]. The Ideal represents what users think performance "can be." The Expected is objective, without an affective dimension, and represents what users think performance "will be." The Deserved has an affective dimension and represents what users feel performance "should be." Lastly, the Minimum Tolerable is what users think the lowest performance "must be."

Based on Miller's work [32], we argue that people likely also have multiple levels of privacy expectations beyond desired and achieved privacy. Therefore, we conceptually distinguish between Expected ("will be") and Deserved ("should be") types of expectation in measuring user expectations for website data practices, and focus on eliciting the Expected ("will be") type to identify mismatches.

We identify mismatches in user expectations regarding website data practices. We study if users expect that a website *will* collect, share or delete data. Prior work has studied mismatches in other types of expectations [13, 16, 26, 33]. To measure expectation, these studies either used an expectation type in the sense of desired preferences (*should*) [33], or they did not clarify the type of expectation [13, 16, 26]. Earp et al. studied Internet users' privacy values and analyzed privacy policies for respective statements [13]. They find

that Internet users' concerns and values are not adequately reflected in privacy policies. Gomez et al. also compared websites' data practices with practices users find concerning [16]. Milne and Bahl examined differences between consumers' and marketers' expectations regarding use of eight information technologies [33]. Liu et al. measured disparity between expected and actual Facebook privacy settings. In contrast to our study on website data practices, Lin et al. studied expectations regarding data practices of mobile apps [24]. Further, their work did not differentiate between different types of expectations, and, while eliciting expectations, did not clarify the type of expectation being elicited.

In contrast to prior work, we propose an approach that facilitates direct comparison of individuals' expectations of what a website's data practices are to the website's actual claims of what they do as stated in their privacy policy.

3. METHODOLOGY

Our goal is to identify mismatches between user privacy expectations regarding website data practices and the practices websites disclose in their privacy policy. We define privacy expectation as what users think a website "will" do or is doing as opposed to what they prefer a website "should" do, which corresponds to Miller's distinction of the Expected and Deserved expectation types [32]. We elicited user expectations for different online scenarios that varied in terms of data practices, website type, and other website characteristics, in order to understand the impact of contextual factors on privacy expectations. We also studied how user characteristics influence expectations. To identify mismatched expectations and unexpected practices, we compared elicited expectations with the data practices described in websites' privacy policies. In the rest of this section, we describe the study design, studied parameters, and the procedure we used to identify and classify mismatched expectations.

3.1 Study Design

To assess the impact of different website scenarios on privacy expectations, we conducted an online study involving 16 websites and 240 participants. We opted for a between-subjects design to prevent fatigue and learning effects, in which we asked participants to answer questions about one website randomly assigned to them. Website type (health, finance, dictionary) and popularity (low, high) were the main independent variables in our study, resulting in a 3x2 design with six conditions. We based website type and popularity on website categories and traffic rankings respectively obtained from Alexa.com [4]. In total, we studied 16 websites, which are listed in Table 1, across three website types (7 Health, 7 Finance, 2 Dictionary). Fifteen participants were assigned to each website, resulting in the following number of participants per condition: 60 in Health-Low, 45 in Health-High, 60 in Finance-Low, 45 in Finance-High, 15 in Dictionary-Low, and 15 in Dictionary-High.

Website	Type	Subtype	Context	Rank
Webmd.com	Health	Reference	Private	107
Medhelp.org	Health	Reference	Private	2,135
Medlineplus.gov	Health	Reference	Government	558,671
Walgreens.com	Health	Pharmacy	Private	315
Bartelldrugs.com	Health	Pharmacy	Private	54,737
Mayoclinic.org	Health	Clinic	Private	297
Clevelandclinic.org	Health	Clinic	Private	2,629
Americanexpress.com	Finance	Credit	Private	76
Discover.com	Finance	Credit	Private	324
Bankofamerica.com	Finance	Bank	Private	33
Woodlandbank.com	Finance	Bank	Private	915,921
Banknd.nd.gov	Finance	Bank	Government	5,267
Paypal.com	Finance	Payment	Private	21
V.me	Finance	Payment	Private	27,289
Merriam-webster.com	Dictionary	–	Private	266
Wordnik.com	Dictionary	–	Private	8,412

Table 1: Websites used in the study (Alexa website rank as of March 10, 2015).

3.1.1 Survey Questionnaire

We designed a questionnaire to measure user expectations for eight collection data practices (4 information types collected with or without account), eight sharing data practices (4 information types shared for core or other purposes), and one deletion data practice. These website practices, listed in Table 2, were treated as 17 dependent variables.

The survey questionnaire consisted of three sections: introduction, main questionnaire and post-questionnaire. Privacy-related questions, which could bias participant responses, were asked in the post-questionnaire. While designing the questionnaire, we used think-aloud and verbal-probing cognitive interviewing techniques [52] in pilot tests with six participants. We tested whether participants understood the questions. We iteratively refined the questionnaire based on participant feedback. We summarize the questionnaire below. The full questionnaire is provided in Appendix B.

At the beginning of the questionnaire, we explained the purpose of the study. We framed the purpose of the study as understanding user opinions about websites rather than their knowledge of data practices, to avoid self-presentation issues associated with knowledge questions [7]. We also did not mention privacy or data practices to avoid biasing participants. After explaining the purpose, we asked whether participants had visited or used the assigned website before.

We instructed the participants to familiarize themselves with the website assigned to them. Since participants may explore websites in different ways, we wanted them to look at what they considered important and did not want to bias their thinking by providing too specific instructions. Based on participant feedback from our in-lab pilot tests, we asked participants to look at the website for 2–3 minutes. Initially, we had instructed the participants to take their time familiarizing themselves with the website. However, after about three minutes of interaction, our in-lab participants were either ready to provide their opinions or were not sure what else to look at. Two participants specifically told us that it would be helpful if we told them how much time they should spend looking at a website. Because the website was opened in a separate browser window, participants could go back to the website at any point during the study.

After participants interacted with the website, we provided definitions of contact, financial, health and current location information. For example, we described contact information as "Examples include (but are not limited to) email address, postal address, phone number, home phone number, etc." Definitions for all information types are provided in Appendix A.

In the main part of the questionnaire, we asked participants about their expectations regarding different website data practices, listed in Table 2. First, we asked them questions about data collection practices in two scenarios: collection without account and collection with account. Before asking questions related to a scenario, we showed scenario descriptions. For instance, for the collection without account scenario, we showed the description "*Imagine that you are browsing [website name] website. You do not have a user account on [website name], that is, you have not registered or created an account on [website name].*" We then asked them about their expectations concerning whether and how the website collects different types of data. These questions were framed as likelihood questions: "*What is the likelihood that [website name] would collect your information in this scenario?*" Note that we framed the questions as "would collect" in order to capture participants' objective expectations, and not what they would prefer. We provided a 4-point scale {Likely, Somewhat likely, Somewhat unlikely, Unlikely} as the response option. We wanted respondents' "best guess" and thus did not provide a neutral or not sure option. We did so because users often do not read privacy policies and decide about data practices of a website based on incomplete information, that is, their best guess. We asked an open-ended question to understand how they thought the website collected their information without having an account on the website. After answering questions about the without account scenario, participants read the scenario description for collection with an account and answered the same questions regarding this scenario.

After collection-related questions, we asked participants questions regarding data sharing practices. We first asked them questions about a scenario where data is shared for core purposes, which we defined as sharing only for the purpose of providing a service that the user requested. We them asked them questions regarding a scenario where data is shared for other purposes, which we defined as a purpose unrelated to providing a service that the user requested. To answer the questions, participants had to understand three concepts. First, what are core purposes for the given website? Second, what are other purposes for the given website? Lastly, with whom could the website possibly share information? To encourage them to think about these concepts, we asked them three open-ended questions before asking questions related to sharing. Concerning the data deletion practice, we asked participants whether they expected that the website would allow them to delete all, some or none of their data.

In the post-questionnaire, we captured different user characteristics in order to study their impact on the participants' privacy expectations. We list these characteristics in Table 3. We ordered the questions based on ease of answering, level of threat, and effect on subsequent answers [7]. First, we asked questions about their *past experiences* with the assigned website including if they had an account on the website, how much they had used the website, familiarity with the website and the website's perceived trustworthiness. Users' past experience may influence their expectations, for example, having an account may expose them to additional parts of a website that may improve their awareness of the website's data practices. Participants then provided demographic information (gender, age, education, occupation) and whether they had a background in computer-related fields, which may indicate an enhanced understanding of online data practices. We also asked for their U.S. state of residence, to assess whether privacy regulation on the state level, e.g., in California, impacts privacy expectations. We further included questions about privacy-protective behavior [37] and their familiarity and knowledge of privacy concepts and privacy-enhancing technologies [21]. We also asked whether participants had negative online experiences [44], as they may expect data practices to be more privacy invasive. Lastly, we included the 10-item IUIPC scale [27] to assess online privacy concerns.

3.1.2 Study Deployment & Demographics

Our study received approval from Carnegie Mellon University's Institutional Review Board. To recruit participants efficiently and rapidly, we used the Amazon Mechanical Turk crowdsourcing platform [5]. Research has shown that the Mechanical Turk sample pool is more diverse than traditional sample pools [40], and that data quality is typically good [8, 40, 41]. In February 2015, we recruited 240 participants. We restricted participation to individuals located in the United States, with at least a 95% approval rate and at least 500 completed tasks on Amazon Mechanical Turk. Participants received $3.50 for completing the study. Each participant was randomly assigned to one of the 16 websites. We implemented our survey on SurveyGizmo. Participants were redirected from Amazon Mechanical Turk to SurveyGizmo to complete the survey. We used a combination of SurveyGizmo and Mechanical Turk features to ensure that participants took the survey only once. We implemented timers to measure how long participants interacted with a website and to measure time spent on survey questions. As instructed, participants, spent on average 1.99 min (SD=2.41, median=1.56) interacting with a website. Statistical analysis did not show a significant impact of the amount of time spent on a website or on the survey questions.

To ensure data quality, we screened for participants that completed the study in less than 10 minutes (pilot tests suggested a 30-minute completion time), and checked whether participants answered two questions about prior experience with the assigned website at the beginning and the end of the survey consistently. All participants passed at least two of three quality criteria.

The 240 participants completed our online survey in 22.5 minutes on average (SD=12.8, median=18.6). The sample was 42% female and 58% male. The average age was 34.4 years (SD=10.3, median=32). The majority (85.3%) had at least some college education and 61.6% reported an Associates, Bachelors or Graduate degree. A fifth of the participants (19.5%) had a college degree or work experience in a computer-related field. The top primary occupations were administrative staff (17.5%), service (14.1%), and business/management/financial (12%).

3.2 Scenario Parameters

We defined multiple scenarios that varied in key parameters, namely data practices and website characteristics. We hypothesized that these parameters may influence privacy expectations and mismatches.

Action	Scenario	Information type
Collection	With account	Contact
		Financial
		Health
		Current location
	Without account	Contact
		Financial
		Health
		Current location
Sharing	For core purpose	Contact
		Financial
		Health
		Current location
	For other purpose	Contact
		Financial
		Health
		Current location
Deletion	–	Personal data

Table 2: Studied data practices.

Website characteristic	
Type	Finance
	Health
	Dictionary
Popularity	More
	Less
Context	Private
	Government
User characteristic	
Demographic: age, gender, education, occupation computer background, state of residence	
Privacy protective behavior	
Familiarity with privacy concepts and tools	
Knowledge of privacy concepts and tools	
Negative online experience	
Online privacy concern	
Experience with website: amount of recent use, has account, familiarity, trust	

Table 3: Studied website and user characteristics.

3.2.1 Data Practices of Interest

We decided to focus on data practices concerning *collection, sharing and deletion of personal information* as prior research has shown that users are especially concerned about surreptitious collection, unauthorized disclosure and wrongful retention of personal information [47]. We considered the collection and sharing of four categories of privacy-sensitive information [1, 19, 23]: *contact information* (e.g., email or postal address), *financial information* (e.g., bank account information, credit card details, or credit history), *health information* (e.g., medical history or health insurance information), and *current location* (e.g., from where a user is accessing the website). The definitions are provided in Appendix A.

We further distinguished between scenarios in which users have or do not have an *account with the website.* Websites typically collect data when users create an account, often explicitly provided by the user. Hence, users may have different expectations depending on whether they have an account or not. In general, users may not be aware of implicit or automated data collection, e.g., of IP addresses and cookies. Websites may use IPs, email addresses and other information to acquire additional data about individuals, such as purchase history or interests, from social media services and data brokers [45].

Similarly, information sharing with third parties, while abundant, is less visible to users. Websites assume to have the users' permission because they are using the website and therefore implicitly consent to its privacy policy. We distinguish between third party sharing for *core purposes*, such as sharing a user's information to provide the requested service (e.g., payment processing or providing contact information to a delivery service), and sharing for unrelated *other purposes*, such as advertising or marketing. In all, we studied 17 data practices summarized in Table 2.

3.2.2 Website Characteristics

To understand whether mismatched privacy expectations vary based on context, we considered three website characteristics: website type, popularity and ownership. *Website type* may influence what information users expect a website to collect [34]. We selected three website categories: finance, health and dictionary. Users may expect finance and health websites to collect sensitive information (health or financial data, respectively). In contrast, users may not expect dictionary websites to collect sensitive information. In the financial category, we included banking, credit card and online payment websites. In the health category, we included pharmacy, health clinic and health reference websites. Website categories were determined using Alexa website categories [4].

Users' expectations may be influenced by their offline interactions with entities affiliated with a website, such as visiting a bank branch or a clinic. Hence, we included websites with *offline interactions* as well as online-only websites in the health and financial categories; dictionary websites were online-only.

Interestingly, popular financial websites have been shown to have more privacy-invasive data practices than less popular ones [12]. Therefore, we studied websites of comparable utility but varying in *popularity*, as determined by their traffic rankings [4].

For a given website type, *government or private ownership* may influence user expectations. Our sample population was limited to the United States, and in the post-Snowden era, people may expect government websites to be more privacy invasive than private websites. Hence, we studied whether user expectations varied between government and privately-owned health and financial websites. Table 3 summarizes the website characteristics that we considered in our model.

3.3 Identifying Mismatched Expectations

To identify mismatched expectations and, thus, unexpected data practices, we compare participants' expectations concerning a specific data practice with the results of our privacy policy analysis with regard to that practice. The infor-

mation about a given website data practice extracted from the website's privacy policy, may be Yes, No, Unclear or Not addressed. We elicited an objective "will" expectation from study participants. They rated their expectation of whether a website *will* engage in a specific data practice on a 4-point scale (Unlikely–1, Somewhat unlikely–2, Somewhat likely–3, Likely–4). These ratings can be interpreted as indications of a positive (Yes) or a negative (No) expectation that can be compared to the policy analysis results. Comparing a website's data practices and users' expectations this way, results in eight potential combinations, as shown in Table 4. For Yes–Yes and No–No, users' expectations match the websites' practices. Yes–No and No–Yes combinations constitute explicit mismatches. For Unclear–Yes, Unclear–No, Not addressed–Yes and Not addressed–No, it is not clear whether expectations are mismatched because the website's policy is unclear or silent on the particular data practice.

It is worth taking a closer look at the implications of the different types of mismatches. Although, both Yes–No and No–Yes are mismatches, they may impact users' perception of privacy violations differently. In the case of Yes–No, the website will collect or share information, but users optimistically expect it not to. Due to lack of awareness that the website shares information, users may decide to use the website. By doing so, they give up data that they do not want to share, resulting in a violation of their privacy. Although the website discloses its data practice in its policy, from a user viewpoint, the practice could be considered surreptitious unless users are appropriately and explicitly made aware of it. When found out, such data practices may damage a company's reputation.

In contrast, in the case of No–Yes, a website will not engage in a collection or sharing practice, but users pessimistically expect it to. As a result, users may have reservations to use the website or some features, which may affect their utility but not their privacy. In such cases, websites should aim to make users aware of the privacy-protective practices to assuage pessimistic expectations.

The number of unclear website data practices can be high, for example, ∼40% of collection data practices in this study are unclear. Hence, it is important to analyze the impact of unclear data practices. Consider the Unclear–Yes case. If the website is really collecting information, then it would be a Yes–Yes match. If the website is not collecting information, then it would be a No–Yes mismatch. The same applies to Unclear–No. As discussed, a Yes–No mismatch, could potentially violate user privacy. Hence, for analysis purposes, we could treat Unclear as a likely Yes. We use a similar approach for Not addressed–Yes and Not addressed–No.

We can similarly analyze mismatches in case of the data deletion practice by considering two types of Yes values, Yes–Full and Yes–Partial, separately. We could also simplify the analysis by combining the two Yes values. In case of deletion, users may use a website if they think that the website allows deletion, whereas for collection and sharing they may not use the website. Hence, in case of deletion, the implications of No–Yes and Yes–No mismatches are reversed.

4. STUDY RESULTS

To identify unexpected practices – those that did not match participants' privacy expectations – we first analyzed the privacy policies of the websites used in our study and then compared them to participants' expectations.

		User:	Yes	No
Website:	Yes		✓	X
	No		X	✓
	Unclear		?	?
	Not addressed		?	?

Table 4: Overview of matched and mismatched expectations. Match (✓) or mismatch (X) between a website's data practice and a user's expectation. If the website's policy is unclear or silent on a practice, it cannot be determined if it matches user expectations (?).

4.1 Website Privacy Policy Analysis

Two annotators, one with legal and another with privacy expertise, independently read each of the 16 privacy policies (cf. Table 1) and extracted the relevant collection, sharing and deletion data practices described earlier. Agreement was generally high, for instance, among the 17 data practices, the highest inter-annotator agreement was κ=1 and lowest agreement was κ=0.718. All disagreements were resolved jointly after initial independent coding. Following an annotation approach similar to Reidenberg et al. [46], annotators coded collection and sharing practices as *yes, no, unclear* or *not addressed*, in order to take ambiguity in the policy language (*unclear*) or silence on a specific practice (*not addressed*) into account. For example, the statement "When you use our Websites, we collect your location using IP address." makes it clear that the website collects location information. However, the statement "We collect the IP address from which you access our Website." mentions collecting IP address but is unclear whether the website collects location information. Collection and sharing practices were analyzed with regard to contact, financial, health and current location information, as well as for two collection contexts (with/without user account) and for two sharing purposes (core/other). Deletion practices were annotated as *full deletion* (websites allows deletion of all user data), *partial deletion* (deletion of only some data), *no deletion*, *unclear*, or *not addressed*. Table 5 shows a sample annotation for Bank of America's privacy policy. Annotating privacy policies is an active area of research, and recent results [6, 53] show the possibility of achieving acceptable level of agreement with semi-automated techniques and non-expert crowdworkers. Such techniques can enable scaling up our approach to large number of websites.

Figure 1 gives an overview of data practices extracted from the privacy policies of the 16 websites (7 financial, 7 health, 2 dictionary) used in our study. It shows the percentage of collection and sharing data practices that are clear, unclear or not addressed in the privacy policies. We find that policies in all three website categories are mostly clear about practices concerning the collection or sharing of contact information, i.e., they make explicit statements about whether they collect or not collect contact information and make clear statements about sharing (dominantly yes for core purposes; no for other purposes).

Data practice	Answer
Collect contact – with account	Yes
Collect contact – without account	Unclear
Collect financial – with account	Yes
Collect financial – without account	No
Collect health – with account	Yes
Collect health – without account	No
Collect location – with account	Unclear
Collect location – without account	Unclear
Share contact – core purpose	Unclear
Share contact – other purpose	Unclear
Share financial – core purpose	Yes
Share financial – other purpose	Yes
Share health – core purpose	Yes
Share health – other purpose	No
Share location – core purpose	Unclear
Share location – other purpose	Unclear
Deletion	No

Table 5: Annotations for the 17 data practices of BankofAmerica.com's privacy policy.

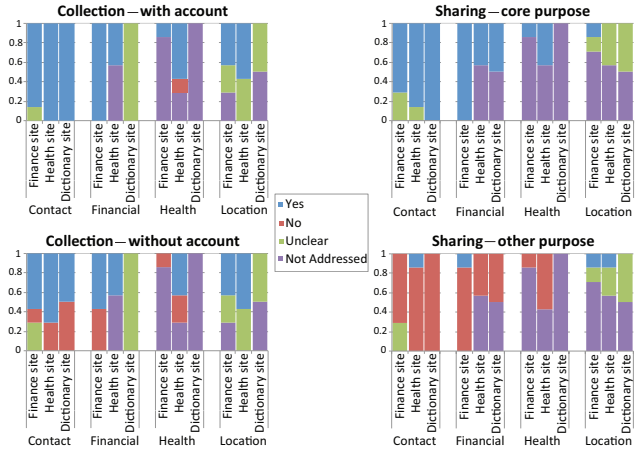


Figure 1: Collection and sharing data practices of the 16 websites used in our study, based on the analysis of the websites' privacy policies.

Not surprisingly, finance websites make explicit statements about collection and sharing of financial information. Note that credit card and online payment finance websites collect financial information even from non-registered users, e.g., when users buy products, but banking websites do not. About half of the health websites' privacy policies also make explicit statements concerning financial information, however, the other half is silent on whether they collect or share financial information. Interestingly, the dictionary websites make statements that leave it unclear if they may collect financial information, but are either explicit or silent on sharing of financial information. Dictionary sites mention processing payments or posting transactions but not explicit collection of financial information.

All dictionary websites and all but one of the financial websites do not address collection or sharing of health information. One of the finance websites, BankofAmerica.com is explicit about collecting health information from registered users and sharing it with third parties for core purposes. It does so via its insurance-related affiliates, which may not be obvious to users. However, all but two of the health websites are explicit about whether they collect health information. Both health clinic websites do not address collection of health information in their website privacy policy, but contain links to additional policies, which may disclose their collection practices. Health websites are less explicit about sharing of health information compared to collection of health information.

About half of the financial and health websites are clear about collection of current location information, but none of the dictionary sites are clear on this aspect. Almost all website privacy policies are unclear or silent on whether they share location information with third parties. Only one finance website explicitly states that it shares user location for core and other purposes. Only one health website explicitly states that it shares user location for other purposes, but it is unclear whether it shares it for core purposes.

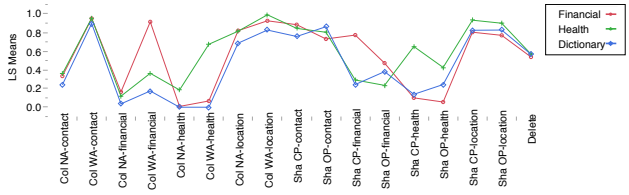
Financial websites are more explicit about deletion data practices compared to health and dictionary websites. Nearly 71% (5) of the financial websites clearly disclose their practice in contrast to 50% (1) of the dictionary websites and 28% (2) of the health websites. However, nearly half of the financial websites (3) do not allow any deletion of data and two only allow partial deletion. In contrast, when clear about the practice, health websites (2) and dictionary websites (1) allow full deletion.

The privacy policy analysis shows that some data practices are common across different website types, whereas others are category-specific or even vary within a category. This suggests that if users would rely on website characteristics to anchor their privacy expectations, these heuristics may lead to mismatches between their expectations and a website's stated data practices.

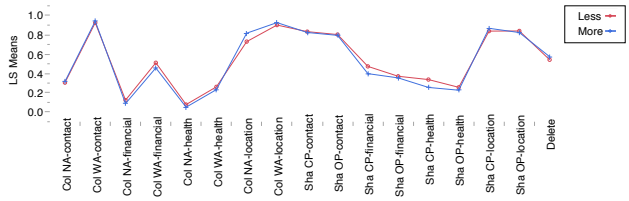
4.2 Impact of Website Characteristics

We find that a website's type has a significant impact on user expectations. This implies that what data practices users expect a website to engage in is influenced by the type of website. We did not find significant differences for popularity or ownership, suggesting they play no or a lesser role in shaping privacy expectations. For example, users expect data practices of BankofAmerica.com, a finance website to be different than those of WebMD.com, a health website. However, they have similar expectations for two finance websites even if one of them is more popular than the other (e.g., in our dataset BankofAmerica.com's popularity rank is 33 and WoodlandBank.com's is 915,921). Similarly, expectations do not differ between privately-owned and government-operated websites. We describe our analysis in more detail in the following.

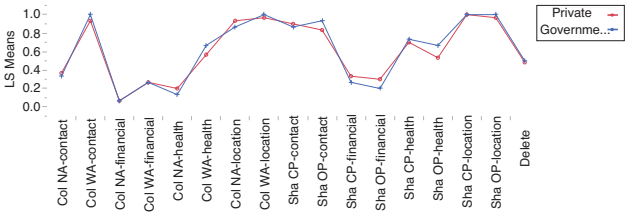
We used a mixed-model ANOVA to analyze the impact of website type and popularity on user expectations. We considered website type (health, finance, dictionary) and popularity (high, low) as nominal between-subjects independent variables. We considered participant expectations concerning the 17 data practices as continuous repeated mea-



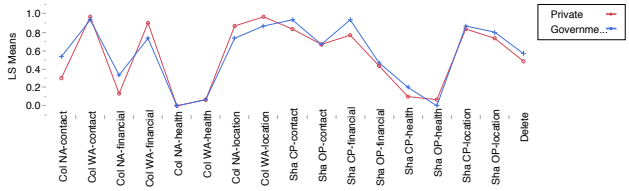
(a) Website type (*sig.*)



(b) Popularity (*n.s.*)



(c) Ownership for health websites (*n.s.*)

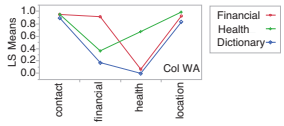


(d) Ownership for finance websites (*n.s.*)

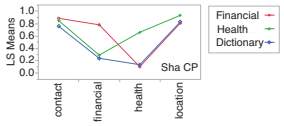
Figure 2: Interaction of website characteristics and user expectations for the 17 data practices. Higher Least Square Means value implies users expect data practice to be more likely (Col: Collection, Sha: Sharing, WA: With Account, NA: No Account, CP: Core Purpose, OP: Other Purpose).

sures dependent variables (DV), which, as a group, measured users' overall expectation. We verified that the group of DVs has an approximate normal distribution with a normal-quantile plot of a linear combination of the individual DV scores. A Shapiro-Wilk W test showed only moderate departure from normality (W=.988, p=.041).

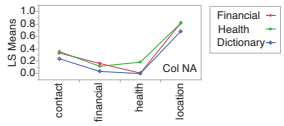
Results showed that interaction of website type and data practices was significant ($F(32.438)$=12.819, p<.0001), see Figure 2a for an interaction plot. This interaction effect suggests that website type impacts what data practices users expect. Compare, for instance, the impact of financial website type on users' expectations concerning collection of financial and health information from registered users (*COL WA-financial*), *COL WA-health*). Higher Least Square Means value implies that users are more likely to expect a data practice. Users expect financial websites to collect financial (high $LSMeans$), but not health data (low $LSMeans$). Figures 2b–2d further show interactions of website popularity



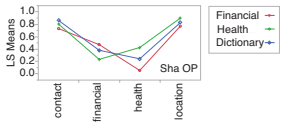
(a) Collection (w. account)



(b) Sharing (core purpose)



(c) Collection (no account)



(d) Sharing (other purpose)

Figure 3: Interaction of website type and expectations for specific data practices. Website type significantly interacts with user expectations for financial and health information. Higher Least Square Means value implies users are more likely to expect a data practice.

and ownership, which were not significant. Note that only the health and finance categories contained government-operated websites, dictionary websites are therefore not shown in Figures 2c and 2d.

We also studied the impact of website type on individual data practices. The distribution of values of individual data practices was non-normal. We treated them as two-level nominal variables and used a χ^2 statistical test. Figure 3a shows what information types participants expect websites to collect from registered users. If $LS\ Means$>0.5, users are likely to expect the data practice. Type of website has a significant impact for expectations of collection of financial ($\chi^2(2,240)$=87.7, p<.0001, R^2=.302) and health information ($\chi^2(2,240)$=105.826, p<.0001, R^2=.3935), but not for collection of contact and current location information. Users expect all types of websites to collect contact and location information when they have an account. However, they expect only financial websites to collect financial data and health websites to collect health data. A financial website collecting health data would lead to a mismatch in expectations. Most financial websites we studied do not collect health data. However, one financial website in our study, BankofAmerica.com, collects health information when users have an account, which violates user expectations.

As shown in Figure 3c, in the without account scenario, participants expect only collection of location information, but for all types of websites. Participants are unlikely to expect websites to collect contact, financial and health data from users without an account. As we will discuss shortly, websites can collect contact and financial data without an account, leading to a mismatch with expectations.

Concerning expectations of data sharing, Figure 3b shows that participants likely expect all types of websites to share contact and current location information for core purposes. Website type has a significant interaction effect for expectations of sharing financial information ($\chi^2(2,240)$=59.175, p<.0001, R^2=.1868) and expectations of sharing health information ($\chi^2(2,240)$ =77.935, p<.0001, R^2=.2642). Participants expect only financial websites to share financial data and health websites to share health data. One financial web-

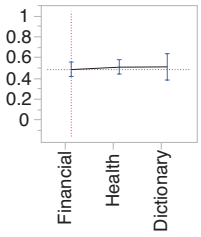


Figure 4: Website type does not impact deletion data practice. LS Means (least square mean) higher value implies users expect data practice to be more likely.

site, BankofAmerica.com, shares health information for core purposes, which violates user expectations.

Figure 3d shows expectations of websites sharing for other purposes. In this case, users expect all types of websites to share contact and location information for other purposes. They do not expect any type of website to share financial or health information for other purposes. Users expecting websites to share contact information for other purposes is interesting because, as we discuss later, most websites do not do so. Lastly, we did not find significant interactions of website type with participants expectations concerning websites' data deletion practices. Participants expected all website types to permit deletion of data, as shown in Figure 4, but this expectation does not match reality.

Further analysis shows that user expectations can vary for individual data types within a larger data type category. For example, for collection of contact information in the with account scenario, participants expected that websites were more likely to collect email address (93.3% participants) than postal address (75%) or phone number (70.8%). Expectations for specific data types can also vary within website sub-categories. For instance, for collection of health information in the with account scenario, participants expected that pharmacy websites were more likely to collect health insurance information than medical history (66.6% vs. 53.3%), but health clinic websites were more likely to collect medical history than health insurance (67.7% vs. 54.8%). Although we could analyze expectations at a finer granularity, identifying mismatches in expectations at finer granularity is problematic because website privacy policies do not typically disclose data practices at such fine granularity. Privacy policies generally discuss data practices at the level of coarse grained categories such as contact information rather than email address or postal address.

4.3 Impact of User Characteristics

We analyzed the effect of multiple user characteristics on participants' data practice expectations. We find that privacy knowledge, privacy concept familiarity, privacy concern, privacy-protective behavior, negative online experience, age, trust in website, website familiarity, whether participant has an account, and recent use have a significant impact on participants' expectations for certain data practices. Other user characteristics elicited in the survey had no statistically significant impact.

For analysis, we considered user characteristics as naturally-occurring, continuous IVs. The DVs were the user expectations for the 17 data practices. Distributions of the individual DVs were non-normal. Therefore, we considered them as two-level nominal variables (Yes, No) and built a nominal logistic regression model for each DV. We assessed internal consistency of summated scale responses using Cronbach's α. For responses to online privacy concern, privacy concept familiarity, privacy knowledge, privacy protective behavior and negative online experience scales, reliability estimates were 0.88, 0.91, 0.63, 0.78, 0.68 respectively. For building regression models, we standardized IV values. To avoid biasing the model due to collinearity of IVs, we computed bivariate non-parametric Spearman rank correlations between IVs and subsequently excluded IVs that had moderate or higher correlation (>0.5). Privacy concept familiarity and privacy-protective behavior were removed from regression models as they correlated with privacy knowledge. Website familiarity and whether the participant has an account were removed because they correlated with the amount of recent use. Our analysis of initial regression models showed that, among demographic variables, only age accounted for a significant amount of variance. Therefore other demographics were removed to improve reliability of the regression models.

As a result, each of the 17 final regression models contained six IVs: privacy knowledge, privacy concern, negative online experience, age, trust in website and recent use. Table 6 lists the user characteristics (IV) and regression models in which the IV was statistically significant in predicting user expectation (DV). Below, we explain the user characteristics (IVs) that can significantly predict user expectations (DVs).

Privacy Knowledge: An individual's privacy knowledge impacts user expectations. Specifically, privacy knowledge can impact if a user expects the collection of health information from unregistered users. An individual with a one unit increase on the privacy knowledge scale is two times more likely to expect that a website will not collect health information without an account. Privacy familiarity and privacy protective behavior correlated with privacy knowledge, and are likely to impact users' expectations in a similar way. Recall that users expect websites, especially non-health websites, to collect health information only when they have an account. If a website did collect health information without an account, there would be a mismatch in expectations.

Privacy Concern: Individuals with higher online privacy concern (IUIPC [27]) expect data practices to be more privacy invasive. Specifically, individuals with one unit increase in online privacy concern are twice as likely to expect that a website will collect current location information when users have an account. They are ∼1.6 times more likely to expect that a website will share contact and current location information for core purposes. Although, most users in our study expect such collection and sharing practices, the segment of users with higher privacy concern are even more likely to expect such practices.

Age: Individuals' age impacts expectations regarding deletion; with one unit increase in age, they are ∼1.8 times more likely to expect that a website will not allow deletion of user data. Older users correctly expect websites not to permit deletion of user data. Hence, the likelihood of mismatch is higher in case of younger users.

Trust in Website: User perception of a website's trustworthiness impacts expectations regarding sharing and deletion data practices. With a one unit increase in trust, individuals are ∼1.7 times more likely to expect that a website will not share health and financial information for other purposes. They are 1.5 times more likely to expect that a website will share location information for core purposes. Lastly, individuals are twice as likely to expect the website to allow deletion of user data. Although, users' expectations based on trust hold for sharing practices, their expectations for deletion does not match reality.

Recent Use: Participants self-reported use of the website in the last 30 days impacts expectations regarding three data practices. With one unit increase in usage, individuals are 1.6 times more likely to expect that a website will not collect current location information from registered users. Individuals are 1.5 times more likely to expect that the website will not share contact information for core purposes. Lastly, individuals are 1.6 times more likely to expect that website will not allow deletion. User expectations are likely to vary similarly based on website familiarity and whether the participant has an account, because both correlated with the amount of recent use. These results confirm our hypothesis that users who have more access to a website have different expectations. However, it is not always true that their expectations are more accurate. For instance, their expectations regarding deletion are more accurate, but expectations regarding sharing are not.

4.4 Matched and Mismatched Expectations

As shown in Figure 5, overall, expected and unexpected data practices varied for different information types, and collection and sharing scenarios. We analyzed mismatches when websites explicitly disclosed their data practices, as well as when websites were unclear or did not address the data practices. When data practices were explicit, we observed three important mismatches. Collection of contact information without an account was mainly a Yes–No mismatch, that is, participants did not expect websites to collect information, but websites did. Similarly, collection of financial information without an account was a Yes–No mismatch. Sharing of contact information for other purposes was also a mismatch, but a No–Yes mismatch, that is, participants pessimistically and incorrectly thought that websites would share their contact information. For the remaining data practices, participants' expectations either predominately matched website practices or the level of match was equal to the level of mismatch.

For the data deletion practice, 32% of participants expected websites to allow full deletion, but only 19% of the analyzed websites allow it. Similarly, 48% expected partial deletion, but only 12% of websites permit it. However, about 20% of the participants thought that websites would not allow deletion of any data and 19% of the websites do not allow deletion of any data. Participants' expectations were similar across the three website types. There is a mismatch in expectations regarding deletion – participants seem to expect websites to allow deletion more than websites actually do.

As we discussed earlier, the number of data practices that are unclear or not addressed in a privacy policy can be high. As shown in Figure 5, websites mostly do not address data

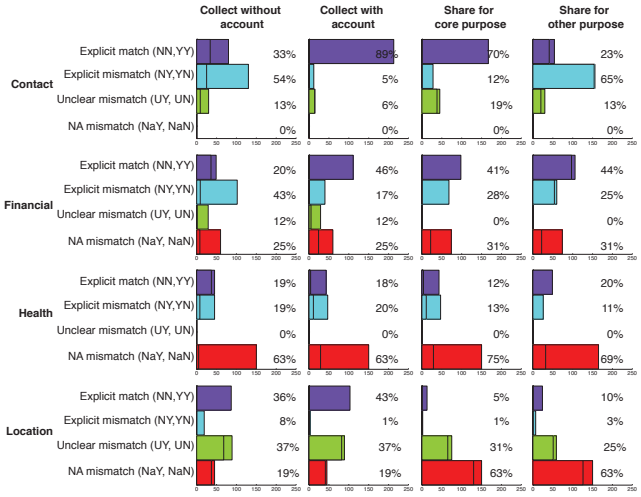


Figure 5: Matches and mismatches in user expectations. Explicit match or mismatch occurs when websites are clear about their data practice. When practice is unclear or not addressed, mismatch is not evident.

practices regarding health information. In contrast, they are mostly unclear or do not address data practices regarding location information. Considering Yes–No mismatches to be more privacy invasive, let us assume that a website engages in a data practice when its disclosure is unclear or not addressed. For health information practices, this results in mainly Yes–No mismatches for all scenarios. However, for location information practices, it results in No–Yes mismatches.

5. DISCUSSION

We identified data practices that do not match user expectations. Our results show that the number of mismatches can be substantial depending on the data practice, and that mismatched expectations vary significantly based on the type of website, as well as user characteristics, such as privacy concern, knowledge, and age. Below, we discuss potential limitations of our study, followed by implications of our results.

5.1 Limitations

We conducted an online study to elicit user expectations. This line of research could benefit from further in-lab studies conducted under more controlled conditions. We compared user expectations with websites' data practices, as disclosed in websites' privacy policies. However, how a website actually handles personal information of their users could potentially be different, but this is difficult to assess in practice.

We recruited participants from Amazon Mechanical Turk. Compared to the general population, they may have higher privacy concern [20], computer knowledge and exposure to privacy-related surveys. Our participants were limited to the United States, and it would be interesting to study expectations of users in other countries or cultures. Nevertheless, our results show that even for potentially more privacy-concerned MTurk participants privacy expectations can be at odds with websites' data practices.

User characteristic (IV)	User expectation (DV)	Model			IV		
		R^2	$\chi^2(6, N{=}240)$	p	Odds(No)	$\chi^2(1, N{=}240)$	p
Privacy knowledge	Collect health info without account	0.10	14.52	0.024	2.09	7.60	0.0058
Privacy concern	Collect location info with account	0.13	13.80	0.0319	0.49	7.22	0.0072
	Share contact info for core purpose	0.09	18.47	0.0052	0.64	5.94	0.0148
	Share location info for core purpose	0.08	15.34	0.0177	0.58	7.67	0.0056
Age	Allow deletion	0.13	30.53	<0.0001	1.77	10.88	0.0010
Trust in website	Share location info for core purpose	0.08	15.34	0.0177	0.65	4.44	0.0352
	Share financial info for other purpose	0.07	21.33	0.0016	1.80	16.82	<0.0001
	Share health info for other purpose	0.05	14.54	0.0241	1.68	1124	0.0008
	Allow deletion	0.13	30.53	<0.0001	0.53	13.64	0.0002
Recent use	Collect location info with account	0.13	13.80	0.0319	1.56	4.01	0.0451
	Share contact info for core purpose	0.09	18.47	0.0052	1.50	6.67	0.0098
	Allow deletion	0.13	30.53	<0.0001	1.56	7.83	0.0051

Table 6: Regression models in which specific user characteristics (IV) significantly impact user expectations (DV). *Odds(No)* indicates, for one unit increase in the IV value, the increase in likelihood that a user will not expect a website to engage in that data practice (*Odds(Yes)*=1 / *Odds(No)*).

We studied collection, sharing and deletion data practices. We asked participants (n=240) if they wanted to know about other data practices; nearly half did not (47.5%). Among the rest, the top three requests were as follows: Participants wanted additional details about sharing (14%). They wanted to know with whom – partners, affiliates and third-parties – their data was being shared. They wanted to know about data security (12%) and how long their data was retained (7%). We plan to extend our research to cover these and other data practices of interest in the future.

We further plan to study more website categories. However, eliciting user expectations for websites with broad or multiple purposes, for example search or social networking websites, is challenging. For example, users may use Google.com for searching, shopping, directions, etc. Along similar lines, it would be interesting to study how accessing multiple websites via a single sign-on impacts expectations. We are studying the impact of additional expectation types, such as the "should" (Ideal) expectation type. Lastly, we are investigating expectations and mismatches in the context of mobile and Internet of Things data practices.

5.2 Highlighting Unexpected Practices

As we discussed earlier, simplified user-facing privacy notices [49] could complement comprehensive privacy policies. Existing simplified privacy notices, for example privacy nutrition labels [22], although an improvement over privacy policies, are themselves too complex. By identifying mismatches in users' privacy expectations, one could selectively highlight or display elements of a privacy nutrition label or other notice format that are most relevant to users. Our results suggest that the number of mismatches is small compared to the total number of website data practices. Thus, likely unexpected data practices should be especially emphasized, and the overall amount of provided privacy information could potentially be reduced. Effectiveness of such highlighting, however, needs to be validated with end users. Different types of mismatches (Yes–No vs. No–Yes) could have different consequences on user privacy, and privacy notices should consider that as well.

Although website operators could themselves generate simplified notices, the low adoption of simplified and standardized notice mechanisms [11] indicates that many website operators may not do so. An alternative approach is for a third-party to highlight unexpected data practices based on mismatched expectations. For example, a browser extension could generate and display a simplified notice [48, 55]. Such a notice could highlight snippets of text from the natural language privacy policy, corresponding to mismatched data practices. Currently third-party browser extensions, such as Ghostery[1] and Privacy Badger,[2] generate and display information regarding online tracking practices. Similarly, a third-party browser extension could display information regarding unexpected data practices. Extensions could use just-in-time notifications or static icons that users can click to gain more information. At installation time, the extension could gather user characteristics such as privacy knowledge, concerns and demographics in order to tailor which practices are emphasized to individual users.

Organizations could also use our approach to obtain a competitive advantage by making their website's data practices and privacy policies easier to understand. In the past, organizations such as Google, have tried to organize information within their policy along dimensions that are important to people, with the intent of making information easier to access. Mismatches in expectations are important, and highlighting them can aid in such efforts. Regulatory agencies such as the Federal Trade Commission work on protecting users' privacy, and mismatched expectations could indicate to them important public policy issues that need attention.

A number of factors are contributing to the growing complexity of website privacy policies. In particular, as websites collect and share more data, policies have to describe more diverse and often more complex data practices. With a growing number of ways to access websites – for example, computers, smart phones, smart cars etc. – policies have to describe data practices that may vary by access mechanisms. Hence, simplified privacy notices that reduce the amount of information to be processed could significantly improve the likelihood of users understanding relevant elements of privacy policies.

[1] www.ghostery.com
[2] www.eff.org/privacybadger

5.3 Generating Simplified Notices

Privacy policies could be potentially simplified or shortened by highlighting data practices that do not match user expectations. For example, consider BankofAmerica.com's privacy policy, which is one of the 16 policies in our study. A full website privacy notice has to include information about all the 17 data practices that we studied. However, for six data practices, user expectations match the website's data practices. Focusing on mismatches, it may be sufficient to highlight those 11 data practices, which is 35% less information. We could further simplify the notice by prioritizing the impact of mismatches. For example, if we determine that Yes–No mismatches are more concerning to users than No–Yes mismatches, the notice could highlight five Yes–No mismatches among the 11 mismatches, which results in 70% less information. This approach could be used in a layered notice approach [49] to determine what practices to include in a high-level summary of the full privacy policy.

Our results indicate that the data practices users expect, as well as respective mismatched expectations, vary significantly by website type. For example, users expect health websites to collect health information, but not finance websites. Therefore, website type could serve as a simple and practical feature to contextualize privacy notices in order to highlight those practices unexpected for the respective website type. Third party tools or browser extensions could further predict, based on website type, which data practices my be unexpected and emphasize or warn about them. Practices that are likely expected for websites of a given type, may not require explicit warnings. For example, in case of the BankofAmerica.com banking website, the extension could signal a mismatch with regard to the website's collection of health information. However, such a warning would not be necessary for health website that collects health information, as most users seems to expect such a practice.

User expectations and mismatches further vary based on user characteristics. Hence, we could personalize privacy notices based on user characteristics. For example, younger users are significantly more likely to expect a website to allow deletion of user data. Hence, when the website does not allow deletion, the likelihood of a mismatch is higher in case of younger users. Thus, privacy decision support tools could highlight a mismatch for younger users only.

Note, that the goal is not to replace or substitute privacy policies, but rather complement them with more targeted notices and tailored warnings to make users aware of those data practices they likely do not expect.

5.4 Semantics and Impact of Mismatches

We discussed mismatches concerning "will" expectations, corresponding to Miller's "Expected" expectation type [32]. We can extend our analysis to additionally include "should" expectations, which are more subjective, as they describe expectations of what would be "Ideal" [32], and are therefore closer to preferences of desired privacy. Users may answer Yes or No to whether a website *should* engage in a data practice. Considering "should" expectations in addition to "will" expectations, would add an additional dimension to the assessment of the implications stemming from matched or mismatched expectations.

For instance, consider when a user's "will" expectation matches the website's data practices (Yes–Yes). When combined with the "should" expectation type, only Yes–Yes–Yes is a perfect match, whereas Yes–Yes–No is a mismatch, i.e., users may expect the practice but prefer it to be different. For example, for data collection, a Yes–Yes–No indicates that a user is correctly aware that a website will collect information, but feels that it should not. The user may continue to use the website due to lack of awareness of other websites that do not collect information. It may also imply market failure due to monopoly or due to all websites in the website category being equally privacy invasive. An example of such market failure may be search engine websites; although users may know that Google's search website collects certain information about them, they may continue to use Google for convenience and utility reasons, despite the availability of privacy-friendly alternatives (e.g., DuckDuckGo.com).

Similarly, in case of a mismatch due to a website engaging in unexpected practices, the "should" expectation type may change the meaning of the mismatch. For example, when a Yes–No mismatch is combined with a "should" expectation. In a Yes–No–No mismatch, users both incorrectly think that a website will not engage in a data practice and feel that it should not. They may decide to use the website and lose data privacy. For Yes–No–Yes, users want the website to engage in a practice, but do not expect it to do so at the moment. For instance, users may want a website to provide personalized services based on their data. In this scenario, users may decide not to use the website and lose utility, but not data privacy.

The examples discussed above demonstrate the importance and potential of distinguishing and capturing the meaning of different expectation types in privacy research. In the case of website privacy notices, by distinguishing between expectation types, we may be able to better identify user needs and display appropriate information. For example, in case of a Yes–Yes–No mismatch, a privacy tool could display alternative websites with more privacy-friendly practices. In case of a Yes–No–Yes mismatch, such a tool could display whether an opt-in option for personalization is available.

Lastly, in addition to the semantics of mismatches, we need to consider which mismatches matter to users. Some mismatches may surprise users, but not really concern them. When designing simplified notices, we could focus on the subset of mismatches that are concerning to users.

6. CONCLUSION

We identified mismatches in user expectations regarding online data practices. Further, we identified factors that impact such mismatches. We believe that emphasizing such mismatches in privacy notices could help users make better privacy decisions. Further, given the small number of mismatches compared to the overall number of data practices, it could be possible to generate simplified user-facing privacy notices as summaries of full privacy policies. Based on the factors that impact mismatches, we identified future research opportunities for contextualizing and personalizing privacy notices and privacy tools to ameliorate the effect of mismatched expectations.

7. ACKNOWLEDGMENTS

This research has been partially funded by the National Science Foundation under grants CNS-1330596 and CNS-1012763. The authors thank Birendra Jha, Ivan Ruchkin and members of the Usable Privacy Policy Project for their useful feedback.

8. REFERENCES

[1] M. S. Ackerman, L. F. Cranor, and J. Reagle. Privacy in e-Commerce: Examining User Scenarios and Privacy Preferences. In *Proc. EC '99*, pages 1–8. ACM, 1999.

[2] A. Acquisti, L. Brandimarte, and G. Loewenstein. Privacy and human behavior in the age of information. *Science*, 347(6221):509–514, Jan. 2015.

[3] I. Altman. The environment and social behavior: Privacy, personal space, territory, and crowding. 1975.

[4] Amazon. Alexa website rankings. `http://www.alexa.com`, 2015.

[5] Amazon. Mechanical turk. `https://www.mturk.com/`, 2015.

[6] J. Bhatia, T. D. Breaux, and F. Schaub. Mining privacy goals from privacy policies using hybridized task recomposition. *ACM Trans. Softw. Eng. Methodol.*, 25(3):22:1–22:24, May 2016.

[7] N. M. Bradburn, S. Sudman, and B. Wansink. *Asking Questions: The Definitive Guide to Questionnaire Design – For Market Research, Political Polls, and Social and Health Questionnaires*. John Wiley & Sons, 2004.

[8] M. Buhrmester, T. Kwang, and S. D. Gosling. Amazon's mechanical turk a new source of inexpensive, yet high-quality, data? *Perspectives on psychological science*, 6(1):3–5, 2011.

[9] F. Cate. The Limits of Notice and Choice. *IEEE Security & Privacy*, 8(2):59–62, Mar. 2010.

[10] Center for Information Policy Leadership. Ten steps to develop a multilayered privacy policy, 2006.

[11] L. F. Cranor. Necessary but Not Sufficient: Standardized Mechanisms for Privacy Notice and Choice. *Journal on Telecommunications and High Technology Law*, 10:273, 2012.

[12] L. F. Cranor, K. Idouchi, P. G. Leon, M. Sleeper, and B. Ur. Are they actually any different? comparing thousands of financial institutions' privacy practices. In *Proc. WEIS 2013*, 2013.

[13] J. B. Earp, A. I. Antón, L. Aiman-Smith, and W. H. Stufflebeam. Examining internet privacy policies within the context of user privacy values. *Transactions on Engineering Management,*, 52(2):227–237, 2005.

[14] Federal Trade Commission. Internet of things: Privacy & security in a connected world. FTC staff report, Jan. 2015.

[15] M. C. Gilly, W. L. Cron, and T. E. Barry. The expectations-performance comparison process: An investigation of expectation types. In *Proc. Conf. Consumer Satisfaction, Dissatisfaction, and Complaining Behavior*, pages 10–16, 1983.

[16] J. Gomez, T. Pinnick, and A. Soltani. Know privacy. Technical report, UC Berkeley School of Information, 2009. `http://knowprivacy.org/report/KnowPrivacy_Final_Report.pdf`.

[17] R. M. Hogarth. *Judgement and Choice: The Psychology of Decision*. John Wiley & Sons, 1987.

[18] C. Jensen and C. Potts. Privacy policies as decision-making tools: An evaluation of online privacy notices. In *Proc. CHI '04*, pages 471–478. ACM, 2004.

[19] A. N. Joinson, U.-D. Reips, T. Buchanan, and C. B. P. Schofield. Privacy, Trust, and Self-Disclosure Online. *Human-Computer Interaction*, 25(1):1–24, Feb. 2010.

[20] R. Kang, S. Brown, L. Dabbish, and S. B. Kiesler. Privacy attitudes of mechanical turk workers and the us public. In *Proc. SOUPS '14*, pages 37–49, 2014.

[21] R. Kang, N. Fruchter, L. Dabbish, and S. Kiesler. "my data just goes everywhere:" user mental models of the internet and implications for privacy and security. In *Proc. SOUPS '15*. USENIX, 2015.

[22] P. G. Kelley, J. Bresee, L. F. Cranor, and R. W. Reeder. A nutrition label for privacy. In *Proc. SOUPS '09*. ACM, 2009.

[23] P. G. Leon, B. Ur, Y. Wang, M. Sleeper, R. Balebako, R. Shay, L. Bauer, M. Christodorescu, and L. F. Cranor. What matters to users? factors that affect users' willingness to share information with online advertisers. In *Proc. SOUPS '13*. ACM, 2013.

[24] J. Lin, S. Amini, J. I. Hong, N. Sadeh, J. Lindqvist, and J. Zhang. Expectation and Purpose: Understanding Users' Mental Models of Mobile App Privacy through Crowdsourcing. In *Proc. UbiComp '12*. ACM, 2012.

[25] F. Liu, R. Ramanath, N. Sadeh, and N. A. Smith. A step towards usable privacy policy: Automatic alignment of privacy statements. In *Proc. COLING 2014*, pages 884–894, 2014.

[26] Y. Liu, K. P. Gummadi, B. Krishnamurthy, and A. Mislove. Analyzing facebook privacy settings: User expectations vs. reality. In *Proc. IMC '11*, pages 61–70. ACM, 2011.

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

[28] S. T. Margulis. On the Status and Contribution of Westin's and Altman's Theories of Privacy. *Journal of Social Issues*, 59(2):411–429, June 2003.

[29] G. Marx. Murky conceptual waters: The public and the private. *Ethics and Information technology*, pages 157–169, 2001.

[30] A. M. McDonald and L. F. Cranor. The cost of reading privacy policies. *ISJLP*, 4, 2008.

[31] A. M. McDonald, R. W. Reeder, P. G. Kelley, and L. F. Cranor. A comparative study of online privacy policies and formats. In *Proc. PETS 2009*, pages 37–55. Springer, 2009.

[32] J. A. Miller. Studying satisfaction, modifying models, eliciting expectations, posing problems, and making meaningful measurements. *Conceptualization and Measurement of Consumer Satisfaction and Dissatisfaction*, pages 72–91, 1977.

[33] G. R. Milne and S. Bahl. Are there differences between consumers' and marketers' privacy expectations? a segment and technology level analysis. *Public Policy & Marketing*, 29(1), 2010.

[34] H. Nissenbaum. Privacy as contextual integrity. *Washington Law Review*, 79:119, 2004.

[35] H. Nissenbaum. *Privacy in Context - Technology, Policy, and the Integrity of Social Life*. Stanford University Press, 2009.

[36] P. A. Norberg, D. R. Horne, and D. A. Horne. The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs*, 41(1):100–126, 2007.

[37] Office of the Australian Information Commissioner. Community attitudes to privacy survey, 2013.

[38] J. S. Olson, J. Grudin, and E. Horvitz. A study of preferences for sharing and privacy. In *Proc. CHI '05*, pages 1985–1988. ACM, 2005.

[39] L. Palen and P. Dourish. Unpacking "privacy" for a networked world. In *Proc. CHI '03*, pages 129–136. ACM, 2003.

[40] G. Paolacci and J. Chandler. Inside the turk understanding mechanical turk as a participant pool. *Current Directions in Psychological Science*, 23(3):184–188, 2014.

[41] G. Paolacci, J. Chandler, and P. G. Ipeirotis. Running experiments on amazon mechanical turk. *Judgment and Decision making*, 5(5):411–419, 2010.

[42] I. Pollach. What's wrong with online privacy policies? *Commun. ACM*, 50(9):103–108, Sept. 2007.

[43] President's Concil of Advisors on Science and Technology. Big data and privacy: A technological perspective. Report to the President, Executive Office of the President, May 2014.

[44] L. Rainie, S. Kiesler, R. Kang, and M. Madden. Anonymity, privacy, and security online. *PEW Research Center*, September 2013.

[45] A. Rao, F. Schaub, and N. Sadeh. What do they know about me? contents and concerns of online behavioral profiles. In *Proc. PASSAT '14*. ASE, 2014.

[46] J. Reidenberg, A. M. McDonald, F. Schaub, N. Sadeh, A. Acquisti, T. Breaux, L. F. Cranor, F. Liu, A. Grannis, J. T. Graves, et al. Disagreeable privacy policies: Mismatches between meaning and users' understanding. *Berkeley Technology Law Journal*, 30(1):39–88, 2015.

[47] J. R. Reidenberg, N. C. Russell, A. J. Callen, S. Qasir, and T. B. Norton. Privacy harms and the effectiveness of the notice and choice framework. *ISJLP*, 11, 2015.

[48] N. Sadeh, A. Acquisti, T. D. Breaux, L. F. Cranor, A. M. McDonald, J. R. Reidenberg, N. A. Smith, F. Liu, N. C. Russell, F. Schaub, et al. The usable privacy policy project. Technical report, CMU-ISR-13-119, Carnegie Mellon University, 2013.

[49] F. Schaub, R. Balebako, A. L. Durity, and L. F. Cranor. A design space for effective privacy notices. In *Proc. SOUPS '15*, pages 1–17. USENIX, 2015.

[50] J. E. Swan and I. F. Trawick. Satisfaction related to predictive vs. desired expectations. *Refining Concepts and Measures of Consumer Satisfaction and Complaining Behavior*, pages 7–12, 1980.

[51] Y. Wang, H. Xia, and Y. Huang. Examining american and chinese internet users' contextual privacy preferences of behavioral advertising. In *Proc. CSCW '16*. ACM, 2016.

[52] G. B. Willis. *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Sage Publications, 2004.

[53] S. Wilson, F. Schaub, R. Ramanath, N. Sadeh, F. Liu, and N. A. Smith. Crowdsourcing annotations for websites' privacy policies: Can it really work? In *WWW*, 2016.

[54] V. A. Zeithaml, L. L. Berry, and A. Parasuraman. The nature and determinants of customer expectations of service. *Academy of Marketing Science*, 21(1):1–12, 1993.

[55] S. Zimmeck and S. M. Bellovin. Privee: An architecture for automatically analyzing web privacy policies. In *Proc. USENIX Security '14*, 2014.

APPENDIX

A. DEFINITION OF INFORMATION TYPES

Contact Information: Examples include (but are not limited to) email address, postal address, phone number, home phone number, etc.

Current location: Current, real-time location of a user accessing the website (city-level or more precise)

Health information: Examples include (but are not limited to) user's medical history, family medical history, user's health insurance information, etc.

Financial information: Examples include (but are not limited to) bank account details, credit/debit card numbers, credit ratings/history etc.

B. SURVEY QUESTIONNAIRE

The complete survey questionnaire is reproduced on the next pages.

[Interview/Survey Questionnaire]

Thank you for your interest in our study.

Your answers are important to us. Please read the instructions carefully so that you can answer our questions as accurately as possible. Take your time in reading and answering the questions.

Peoples' opinions about websites may or may not vary depending on the type of website (news, health, finance etc.) and past experience (not heard of website, heard of, not visited, visited etc.)
While answering questions about a website, **think about your interactions only with the website**. Your interactions could be through a computer, mobile phone or other device. Ignore any interactions with mobile apps, physical stores, businesses or other websites related to the website.
For each website listed below, select the option that best indicates your answer.

	I have not heard of it	I have heard of it, but not visited it	I have visited it, but not in the last 3 months	I have visited it in the last 3 months	Don't know/Not sure
[website]	○	○	○	○	○

I would like to understand your opinions regarding Internet websites. For any question, it is okay to say that you don't know the answer. If you are guessing an answer, please say so. It would be very helpful, if you explain your reasoning behind your answers.

[For each website assigned to a participant, ask the following questions]

Now, I would like your opinions regarding [website name] website. Please interact with the website (provide URL) for 2-3 minutes and get familiar with it. Please let me know when you are ready to provide your opinions.

1. As far as you can recall, have you used any websites similar to [website name]?
   Yes (please specify) / No

[Omit questions 2 and 3 if the participant has not used the website]

2. I would like you to think about the last time you visited [website name]. As far as you can recall, what did you do on the website?
3. What other things have you done on this website?

To help you answer my questions, I will explain a few terms. Please use this handout to follow along. You can refer back to the handout at any time.

[Provide handout containing definitions for contact/health/financial/current location information]
[Read definitions for contact/health/financial/current location information]

4. Consider the following scenario to answer the next question.

   Imagine that you are browsing [website name] website. You **do not have a user account** on [website name], that is, you have not registered or created an account on [website name].

   What is the likelihood that [website name] would **collect your information** in this scenario? Each row in the table below, lists a specific type of information about you. For each information type, select the likelihood that [website name] would collect that information in the scenario described above.

		Likely	Somewhat likely	Somewhat unlikely	Unlikely
Collects your **Contact information**	Email address	○	○	○	○
	Postal address	○	○	○	○
	Phone number	○	○	○	○
	Other Please specify	○	○	○	○

		Likely	Somewhat likely	Somewhat unlikely	Unlikely
Collects your **Health information**	Medical history	○	○	○	○
	Health insurance information	○	○	○	○
	Other Please specify	○	○	○	○

		Likely	Somewhat likely	Somewhat unlikely	Unlikely
Collects your **Financial information**	Bank account details	O	O	O	O
	Credit or debit card number	O	O	O	O
	Credit rating	O	O	O	O
	Other Please specify	O	O	O	O

		Likely	Somewhat likely	Somewhat unlikely	Unlikely
Collects your **Location information**	Current location (city-level or more precise)	O	O	O	O

5. What leads you to think that [website name] would collect your information when you do not have an account? Please explain.
6. Now, consider an alternate scenario.

   Imagine that **you have a user account** on [website name], and you **have logged in** to your account while browsing [website name].

   What is the likelihood that [website name] would **collect your information** in this scenario?
   Each row in the table below, lists a specific type of information about you. For each information type, select the likelihood that [website name] would collect that information in the scenario just described.

		Likely	Somewhat likely	Somewhat unlikely	Unlikely
Collects your **Contact information**	Email address	O	O	O	O
	Postal address	O	O	O	O
	Phone number	O	O	O	O
	Other Please specify	O	O	O	O

		Likely	Somewhat likely	Somewhat unlikely	Unlikely
Collects your **Health information**	Medical history	O	O	O	O
	Health insurance information	O	O	O	O
	Other Please specify	O	O	O	O

		Likely	Somewhat likely	Somewhat unlikely	Unlikely
Collects your **Financial information**	Bank account details	O	O	O	O
	Credit or debit card number	O	O	O	O
	Credit rating	O	O	O	O
	Other Please specify	O	O	O	O

		Likely	Somewhat likely	Somewhat unlikely	Unlikely
Collects your **Location information**	Current location (city-level or more precise)	O	O	O	O

Thank you. As you may know, companies that own websites may handle information collected on websites in different ways. Some companies share the collected information with other companies, and some companies do not share. Companies may have to share your information in order to provide you a service that you requested on a website.

7. In your opinion, what services can you get from [website name]? Please explain.

8. In order to provide you services, [website name] may have to share your information with other companies. In your opinion, what are those companies, if at all any? Please explain.

9. A website may share your information for purposes unrelated to providing you a service that you requested from the website. What do you think are such unrelated purposes for which [website] can share your information? Please explain.

Before sharing your information, companies may or may not ask for your permission. Some companies assume that the permission is implied because you are using the website. Other companies may explicitly ask you for permission before sharing information, for example, via an explicit written or oral consent.

10. Consider the following scenario to answer the next question.

Imagine that [website name] is sharing your information with another company, but **only for the purpose of providing you a service you requested** on [website name]. Since [website name] has to provide you a service that you requested, [website name] assumes that it has your permission to share information, that is, your permission is implied. [Website name] will share only the information required to provide you the requested service.

What is the likelihood that [website name] would **share your information** with your implied permission in this scenario?

		Likely	Somewhat likely	Somewhat unlikely	Unlikely
Shares your **Contact information**	Email address	O	O	O	O
	Postal address	O	O	O	O
	Phone number	O	O	O	O
	Other Please specify	O	O	O	O
		Likely	Somewhat likely	Somewhat unlikely	Unlikely
Shares your **Health information**	Medical history	O	O	O	O
	Health insurance information	O	O	O	O
	Other Please specify	O	O	O	O
		Likely	Somewhat likely	Somewhat unlikely	Unlikely
Shares your **Financial information**	Bank account details	O	O	O	O
	Credit or debit card number	O	O	O	O
	Credit rating	O	O	O	O
	Other Please specify	O	O	O	O
		Likely	Somewhat likely	Somewhat unlikely	Unlikely
Shares your **Location information**	Current location (city-level or more precise)	O	O	O	O

11. Consider the following alternate scenario to answer the next question.

Imagine that [website name] is sharing your information with another company for **a purpose unrelated to providing you a service you requested**. Since you are using [website name], it assumes that it has your permission, that is implied permission, to share your information for any purpose.

What is the likelihood that [website name] would **share your information** with your implied permission in this scenario?

		Likely	Somewhat likely	Somewhat unlikely	Unlikely
Shares your **Contact information**	Email address	O	O	O	O
	Postal address	O	O	O	O
	Phone number	O	O	O	O
	Other Please specify	O	O	O	O
		Likely	Somewhat likely	Somewhat unlikely	Unlikely
Shares your **Health information**	Medical history	O	O	O	O
	Health insurance information	O	O	O	O
	Other Please specify	O	O	O	O
		Likely	Somewhat	Somewhat	Unlikely

			likely	unlikely	
Shares your **Financial information**	Bank account details	O	O	O	O
	Credit or debit card number	O	O	O	O
	Credit rating	O	O	O	O
	Other Please specify	O	O	O	O
		Likely	Somewhat likely	Somewhat unlikely	Unlikely
Shares your **Location information**	Current location (city-level or more precise)	O	O	O	O

Thank you. As you may know, websites may allow users to delete or remove their data from the website e.g. by closing an account. Allowing users to edit or modify their data is not same as deleting data.

12. Do you think that [website name] would allow you to delete your personal data?
    - O Yes, it will allow me to delete all of my data
    - O Yes, but it will only allow me delete some of my data
    - O No, it will not allow me to delete my data

13. We discussed data practices such as collection and sharing of four types of information, and also deletion of information. What else would you like to know about [website name]?

[End of the interview]

Thank you. That was all I had to discuss. Would you care to add anything?

Thank you. Please take a few minutes to fill out the following questionnaire. That would be the end of our study.

Different users may have different opinions regarding websites. To help us understand how user opinions vary, please answer the following questions.

Please tell us about your experience with <u>[website name]</u> website.

As far as you know, do you have a user account on the website?
( ) Yes, I have an account
( ) No, I don't have an account
( ) Not sure

How many times have you visited the website in the last 30 days? Exclude the visit as part of today's study.
(Please specify a number equal to or greater than zero) _____

In your opinion, how much have you used the website in the last 30 days? Exclude use as part of today's study.

1 – Not at all	2 – Very little	3 - Somewhat	4 – Quite a bit	5 – A great deal
O	O	O	O	O

Do you know someone else who uses the website?
( ) Yes, I know someone
( ) No, I don't know anyone
( ) Not sure

In your opinion, how familiar are you with the website?

1 – Not at all	2 – Slightly	3 - Somewhat	4 – Moderately	5 – Extremely
O	O	O	O	O

In your opinion, how trustworthy is the website?

1 – Not at all	2 – Slightly	3 - Somewhat	4 – Moderately	5 – Extremely
O	O	O	O	O

As far as you know, do you have a user account on a website similar to [website name]?
( ) Yes, I have an account
( ) No, I don't have an account
( ) Not sure

Please tell us about your background.

What is your year of birth (4-digit, yyyy format)?
_____

What is your gender?
( ) Male ( ) Female ( ) Decline to answer

Which of the following best describes your primary occupation?
[List of occupations here]

Which of the following best describes your highest achieved education level?
[List of education levels here]

Do you have a college degree or work experience in computer science, software development, web development or similar computer-related fields?
( ) Yes ( ) No ( ) Decline to answer

Do you currently work or reside in the state of California?
( ) Yes ( ) No ( ) Decline to answer

While using the Internet, have you ever done any of the following things? Please check all that apply.
[ ] Used a temporary username or email address
[ ] Used a fake name or untraceable username
[ ] Given inaccurate or misleading information about yourself
[ ] Set your browser to disable or turn off cookies
[ ] Cleared cookies and browser history
[ ] Used a service that allows you to browse the web anonymously, such as a proxy server, Tor software, or a virtual private network
[ ] Encrypted your communications
[ ] Decided not to use a website because they asked for your real name
[ ] Deleted or edited something you posted in the past
[ ] Asked someone to remove something that was posted about you online
[ ] Used a public computer to browse anonymously

How would you rate your familiarity with the following concepts or tools?

	I've never heard of this.	I've heard of 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 this works.	I know very well how this works.
IP address	O	O	O	O	O
Cookie	O	O	O	O	O
Incognito mode / private browsing mode in browsers	O	O	O	O	O
Encryption	O	O	O	O	O
Proxy server	O	O	O	O	O
Secure Sockets Layer (SSL)	O	O	O	O	O
Tor	O	O	O	O	O
Virtual Private Network (VPN)	O	O	O	O	O
Privacy settings	O	O	O	O	O

Please indicate whether you think each statement is true or false. Please select "I'm not sure" if you don't know the answer.

	True	False	I'm not sure
Incognito mode / private browsing mode in browsers prevents websites from collecting information about you.	O	O	O
Website cookies can store users' logins and passwords in your web browser.	O	O	O
Tor can be used to hide the source of a network request from the destination	O	O	O
A VPN is the same as a Proxy server.	O	O	O
IP addresses can always uniquely identify your computer.	O	O	O
HTTPS is standard HTTP with SSL to preserve the confidentiality of network traffic.	O	O	O
A request coming from a proxy server cannot be tracked to the original source.	O	O	O

In order to protect your personal information, how often have you done the following?

Check that a website is secure before providing personal information.

1 - Never	2 - Rarely	3 - Sometimes	4 - Often	5 - Always
O	O	O	O	O

Ask public or private sector organizations why they need your information.

1 - Never	2 - Rarely	3 - Sometimes	4 - Often	5 - Always
O	O	O	O	O

Read privacy policies and notifications before providing personal information.

1 - Never	2 - Rarely	3 - Sometimes	4 - Often	5 - Always
O	O	O	O	O

As far as you know, have you ever had any of these bad experiences as a result of your online activities?

	Yes	No
Something happened online that led you into physical danger	O	O
Been stalked or harassed online (sexually harassed, physically threatened)	O	O
Got into trouble with local authorities, or government because of your online activities	O	O
Experienced trouble in a relationship between you and a family member or a friend because of something you posted online	O	O
Had your personal information leaked by a company	O	O
Lost a job opportunity or educational opportunity because of something you posted online or someone posted about you online	O	O
Had your reputation damaged because of something that happened online	O	O
Been the victim of an online scam and lost money	O	O
Had important personal information stolen such as your Social Security Number, your credit card, or bank account information	O	O
Something else bad happened (please explain)	O	O

You are almost done. Please share your opinion about Internet consumer experience.
Please indicate how much you agree or disagree with the following statements:

Consumer online privacy is really a matter of consumers' right to exercise control and autonomy over decisions about how their information *is collected, used, and shared.*

Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

Consumer control of personal information lies at the heart of consumer privacy.

Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

I believe that online privacy is invaded when control is lost or unwillingly reduced as a result of a marketing transaction.

Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

Companies seeking information online should disclose the way the data are collected, processed, and used.

Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

A good consumer online privacy policy should have a clear and conspicuous disclosure.

Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

It is very important to me that I am aware and knowledgeable about how my personal information will be used.

Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

It usually bothers me when online companies ask me for personal information.

Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

When online companies ask me for personal information, I sometimes think twice before providing it.

Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

It bothers me to give personal information to so many online companies.

Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

I'm concerned that online companies are collecting too much personal information about me.

Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

Thank you for participating in our study.