As Advertised? Understanding the Impact of Influencer VPN Ads

Omer Akgul◇⌣   Richard Roberts◇   Emma Shroyer◇   Dave Levin◇   Michelle L. Mazurek◇

◇*University of Maryland*   ⌣*Carnegie Mellon University*

Abstract

Influencer VPN ads (sponsored segments) on YouTube often disseminate misleading information about both VPNs, and security & privacy more broadly. However, it remains unclear how (or whether) these ads affect users' perceptions and knowledge about VPNs. In this work, we explore the relationship between YouTube VPN ad exposure and users' mental models of VPNs, security, and privacy. We use a novel VPN ad detection model to calculate the ad exposure of 217 participants via their YouTube watch histories, and we develop scales to characterize their mental models in relation to claims commonly made in VPN ads. Through (pre-registered) regression-based analysis, we find that exposure to VPN ads is significantly correlated with familiarity with VPN brands and increased belief in (hyperbolic) threats. While not specific to VPNs, these threats are often discussed in VPN ads. In contrast, although many participants agree with both factual and misleading mental models of VPNs that often appear in ads, we find no significant correlation between exposure to VPN ads and these mental models. These findings suggest that, if VPN ads do impact mental models, then it is predominantly emotional (i.e., threat perceptions) rather than technical.

1   Introduction

Do ads actually change the way people think? This simple question has been central to a vast body of academic literature in the fields of marketing and advertising, leading to myriad findings about how the frequency and messaging of ads can impact a viewer's brand awareness and willingness to purchase a given product (e.g., [1–11]). Similarly, public health experts have studied the impact that ads can have on health choices, such as how prescription drug ads influence patients' requests for the advertised drugs [12].

Recently this question has also become critical to the security community, with the proliferation of advertising for security & privacy tools. Influencer VPN ads[1] on YouTube, in

particular, have recently become near-ubiquitous. These ads tend to include a significant amount of "educational" content, informing viewers of potential attacks and the defenses that VPNs provide. We had previously studied the content of a random sample of influencer VPN ads, finding extensive instances of misleading or false claims [13], such as that using a VPN means "you won't ever have to worry about anything on the internet again" [14]. Unfortunately, this misinformation is not limited to a small audience; we previously estimated that influencer VPN ads have received over *4.5 billion views* on YouTube alone. This previous study takes a first step toward understanding these ads, but provides no evidence on user impact. Is our speculation on internet threat and VPN mental model influence correct?

In this paper, we look beyond the basic question of whether VPN ads influence viewers to purchase products and ask: what impact do they have on viewers' security & privacy mental models? To the best of our knowledge, we are the first to seek to directly measure the impact of *any* entertainment media on viewers' security & privacy mental models. Prior studies investigating mental models [15–20] often report media depictions as a top source of security & privacy information [16, 21]. However, these findings are based on users' self-reported data, which, while valuable, is subject to limitations in what participants can (and are willing to) recall.

We introduce a novel, multi-stage user study design in which we gather not only participants' mental models, but also *their entire YouTube viewing history*. In total, we collected this data from 217 YouTube users. Leveraging this rich data, we expand on self-reported data by *directly measuring* real-world ad exposure and investigate its correlation with security & privacy mental models.

We find significant correlations between exposure to VPN ads and both VPN brand familiarity and belief in (hyperbolic) threats. However, we find no such relationship between exposure to VPN ads and belief in specific factual or misleading claims about VPN capabilities that commonly appear in VPN

---

[1]Unlike interstitial ads (delivered by YouTube before, during, and after

videos), influencer ads are part of the video, typically produced by influencers, integrated into the content, and without an explicit business tie to YouTube.

ads. Qualitative analysis of our data suggests that participants hold similar, but not identical, mental models to those disseminated through VPN ads.

These findings suggest two possible (not mutually exclusive) interpretations: that VPN ads have an impact on influencing users threat models, or that advertisers (VPN companies, the YouTubers, or the YouTube content delivery algorithms) intentionally invoke threats that resonate with and thereby reinforce models the viewers already have. In either case, the results indicate that emotion plays a larger role than the technical details in VPN ads [13, 22].

Contributions   We make the following contributions:

- We introduce a novel user study design to collect VPN ad-relevant mental models and measure users' VPN ad exposure using entire YouTube viewing histories. We collect a dataset of 217 YouTube users using this method.

- We demonstrate, for the first time, real-world (not merely self-reported) correlation between entertainment media and users' security mental models.

- We show that YouTube users are extensively exposed to VPN ads. This exposure strongly correlates with brand familiarity and belief in hyperbolic threats advertised.

2   Background and Related Work

In this section we summarize related work on VPN mental models and use cases, the use of YouTube as an information source for users, and provide background on modeling the influence of advertisements.

VPN uses and mental models   User mental models about VPNs have received attention from usable security researchers, motivated by increasing numbers of commercial VPNs and their consumers. Mental models are typically defined as users' internal representations of reality [23]. We use a more constrained definition, common in security & privacy work: a user's comprehension of how a system operates, including its inputs, outputs, and the expected effects. Namara et al. found that tech-savvy VPN users tended to use VPNs for non-privacy-related use cases like accessing geo-locked content [20]. However, the same study found that users motivated by privacy tend to use VPNs for *longer*. Ramesh et al. [24] found contradictory results investigating another large (likely) tech-savvy population: the majority of their participants reported using VPNs for security reasons. They also find that users rely on recommendation websites when choosing *which* VPN to use, but it remains unclear how users *initially* learn about VPNs. Dutkowska-Zuk et al. find greater concern with content access and less concern with security & privacy, with students compared to a crowd-sourced general population [25]. Again contradicting prior work, they find that censorship avoidance is a major motivation for U.S. users.

Story et al. found misaligned VPN mental models to be common in a demographically representative online sample [26]. Researchers have noted users adopting VPNs to increase security & privacy while on public networks, as well as to prevent hacks or password leaks [27]. Researchers have found that roughly 6% of Tor users use VPNs in conjunction, potentially indicating the influence of popular media [28]. While using VPNs to access Tor might provide benefits under narrow threat models (e.g., preferring the ISP to know a VPN connection is happening instead of Tor, the VPN is trustable), in many cases it provides no benefits and, depending on the trustworthiness of VPN [29–31], may even harm privacy [28]. Our prior work reported YouTubers conveying similar (dubious) ideas about VPN features and benefits [13], hinting at the possible influence of VPN ads on mental models. Note that VPNs are not a panacea for security & privacy; prior work has found some VPNs that range from misconfigured, to dishonest and even malicious [29–35].

Our work builds on this body of work by investigating the (potential) *source* of these misaligned mental models. Some of the same themes from prior work are replicated in our data.

Information on YouTube   Researchers have found YouTube to be the top online resource for users seeking information about multiple topics [36] including topics that users otherwise know nothing about [37]. This makes YouTube an interesting object of study for security & privacy education, as prior security research has found that some users build their mental models of security & privacy using (online) media [15–17] and ads [21].

We had previously speculated about the impact YouTube might have on mental models through the context of VPN ads [13]. Through a random sample of VPN ads, we had found that these ads are a potential source of security & privacy education that reaches a broad audience. Troublingly, we had found that some influencer VPN ads contain vague, misleading, and/or false information about both VPNs and internet threats. This poses a risk to viewers by potentially distorting viewers' mental models and compromising their security behaviors. These findings may be amplified by the influencer effect: influencer ads are both powerful and cost-effective ways to influence consumers [38–41].

Motivated by these findings, our research aims to directly investigate the impact of VPN ads, by looking for correlations between user mental models and exposure to prominent advertising themes we previously reported [13].

Misleading ads   At a broader level, ads that are inappropriate, deceptive, or otherwise manipulative are commonplace on the internet. Researchers have uncovered such ads in multiple contexts: on social media targetting minorities [42, 43], on news websites before elections [44], potentially illegally on childrens websites [45], and more.

Risk Communication   According to the human-in-the-loop framework, users act as receivers of communications

(e.g., warnings) which may influence their behavior [46]. An extensive body of literature has investigated different ways to effectively convey risks and threats to users. Notably, communication efforts are often hindered by user habituation [47] Bravo-Lillo et al. demonstrated that attractors, or attention capturing UI elements, could reduce warning habituation [47]. In 2017, Albayram et al. showed that fear appeals were effective at inducing behavior change [48]. In a replication of that study, Qahtani et al. found that fear appeals were especially effective when targeting the shared fears of a specific population [49]. Effective risk communication can even influence purchasing decisions; researchers showed that users were willing to pay a premium for privacy-preserving IoT products given appropriate risk communication [50, 51].

We extend this body of work by investigating VPN ads as risk communications that frequently include fear appeals, are often targeted to the audience of a specific YouTube channel, and are inherently designed to persuade users to purchase a brand's product. In contrast with prior work, we investigate the relationship between natural exposure to VPN ads in the wild and users' mental models of security threats.

Modeling the Influence of Ads Modeling the effects of repeated ad exposure on attitudes and behaviors is a long-standing research problem [2]. Despite an extensive body of research, there is no consensus on the exact relationship; studies have suggested linear [1, 2], quadratic [1–4], cubic [1], radical [5], logarithmic [6–8], and various discontinuous relationships [9–11]. Complicating the issue, these models are often dependent on several covariates such as brand familiarity, product category, and advertising medium, and various advertising traits.

A large body of work argues the existence of "wear-in/wear-out" effects, where repeated exposure to an ad first increases, then decreases, attitudes towards products and use intent [2, 11, 52]. Most model this effect with quadratic terms [1–4] though some model the phenomenon as an *inverted U* [53]. Following this body of work and based on preliminary data exploration, we explore whether the relationship between repeated exposure to VPN ads and user perceptions follow this inverted U relationship.

Although there has been extensive research on traditional ads (TV ads, banners, etc.), empirical research on influencer ads is limited. Further, to the best of our knowledge, the impact of influencer ads for security & privacy products remains entirely unexplored. We select our analysis methods knowing that prior work in advertising relavent, but do not expect these methods to explain our data perfectly.

3 Methods

We hypothesize a relationship between a user's exposure to influencer VPN ads and their belief in what ads convey. Specifically, we design our study to answer these key questions:

0. Does VPN ad exposure correlate with VPN brand familiarity?[2]

1. Does VPN ad exposure correlate with VPN mental models?

2. Does VPN ad exposure correlate with internet threat mental models?

3. What are the most common threats people believe VPNs protect against? Do beliefs mirror prior works' findings?

This section describes our study design, independent and dependent variable definitions, and analysis methodology. We pre-registered our analysis plan[3] after our preliminary studies but before the final data collection; our final analysis did not deviate from our original plan.

3.1 Study design

Our study comprised three stages: (1) an initial screener, (2) a tutorial stage, and (3) a final questionnaire stage. Informed by our extensive piloting and preliminary studies (§3.2), we designed the staging of questions with multiple optimization goals in mind: minimize fatigue and dropout, while maximizing data quality and comfort in the study. Table 1 summarizes our major data sources and in which stage we collected them. The screener and tutorial stages were separated by hours to two days, while the tutorial and final questionnaire stages were separated by however long it took participants to obtain the relevant data (in practice, minutes to 15 days).

Participants were recruited from Prolific. We aimed for a minimum of $12/hour (well above U.S. federal minimum wage and consistent with Prolific recommendations[4]) for time spent on tutorials and surveys. Because some tutorials were shorter than others, in practice some participants received a higher hourly rate (5 mins max, $1.00 per tutorial). We paid for data separately ($5.25), totaling $9.52–$10.52 per completed submission. Further, we gave bonuses to participants who spent time resolving technical issues and paid data compensation to participants who made an effort even if their data upload was ultimately unsuccessful.

Before each stage, participants were given an overview of the the current and future stages. All procedures (detailed below) were approved by the University of Maryland Institutional Review Board (IRB).

3.1.1 Screener stage

We recruited from gender balanced Prolific users who were 18 years or older, lived in the US, and used YouTube at least once a month. The study was advertised vaguely as "Internet Perceptions Study," and consent was obtained just for

---

[2]A baseline to validate our method. Ads increase brand familiarity [2]
[3]https://aspredicted.org/rk8xe.pdf
[4]https://researcher-help.prolific.com/hc/en-gb/articles/4407695146002-Prolific-s-payment-principles

Data	Explanation	Collection stage
Through questionnaires:		
Privacy sensitivity	IUIPC-8 scores.	Screener/tutorial
Hardware configurations	Computer and mobile phone OS	Screener
VPN mental models	Threat, misleading VPN, factual VPN, all all VPN models	Final
Brand familiarity	How familiar participants are with certain brands	Screener/tutorial
Through tutorials:		
YouTube watch histories	Google Takeout YouTube histories, later augmented with video subtitles and details.	Tutorial/final
App download histories	Android or Apple app download histories. Searched for VPN applications.	Tutorial/final
Computer program list	List of programs on Windows or MacOS device. Searched for VPN applications.	Tutorial

Table 1: Major data sources collected from participants, grouped by collection technique (self-report vs not).

the screener intentionally to keep the real purpose vague (described later in this stage). A second formal consent with details of the study was obtained before the tutorial stage. We chose this two-tiered consent approach to (1) alleviate selection bias and (2) to be able to measure the privacy sensitivities of participants who ultimately could or would not continue.

After the screener consent, participants were asked questions on their device and YouTube usage. A small group (n=138) of participants were also asked to complete the tutorial stage's questionnaire (this group skipped this questionnaire later). We use this group to compare the privacy sensitivities of participants who did and did not complete the study. The group was limited since was impractical and statistically unnecessary to obtain these from every screener participant.

Finally, we described the full study to participants and asked if they were willing to share the data that we needed for the study. Participants were able to select however many data sources they were willing to share, but it was clear that those who did not give all three sources would not continue past the screener. Though not required by our ethics committee, this question is essential to the consent structure of our study.

We admitted any participant to the tutorial stage who met the following criteria: (1) reported to use an Android or iOS primary mobile device, (2) reported to use a Windows or Mac primary computer, (3) watch at least three videos a week on average,[5] (4) "frequently" or "always" be signed in when watching YouTube, (5) be "rarely" or "never" [6] not watching when others were watching YouTube videos through their accounts (this filters out children who use parents' devices, communal TVs signed into one account, etc.), (6) have had their YouTube history on for at least the past year, and (7) consented to sharing their data (Android or Apple account app download histories, installed program list on their primary computer, and YouTube histories). Participants who did not qualify were paid for their time.

At this stage, participants' hardware configurations and data-sharing statuses were recorded automatically on a data

collection server we instrumented with custom software. We describe the data collection infrastructure in §3.1.2. The full set of questions is in the extended paper (Appendix A).

3.1.2 Tutorial stage

Participants who chose to continue with the study were asked for consent once again with full study details. Once consent was obtained, participants first completed a questionnaire consisting of IUIPC-8, questions on VPN use (have they used a VPN, what purpose, and the brand names), their familiarity with the most popular VPN brands (determined by our exploratory data and asked on Likert-type options: "Not at all familiar (1)"–"Very familiar (7)"), and an open-ended question asking participants to list the top two threats VPNs protect against to the best of their knowledge. Those who finished the questionnaire were redirected to the tutorials.

The tutorial stage consisted of multiple sub-tutorials guiding participants to provide data or start the exports necessary to provide data later in the final stage. The system used hardware configurations of participants (collected during the screener) and dynamically populated the necessary tutorials. Participants, thus, followed multiple flows. Sub-tutorials were combined where possible (e.g., Android app download and YouTube histories). Generally, participants were asked to: (1) Start the export of their YouTube histories via Google Takeout,[7] (2) Start the export of their app download histories (Google Takeout or privacy.apple.com), and (3) export the list of programs on their primary computer.[8]

Once tutorials were over, participants were asked to upload any data that had already finished exporting, frequently the case for YouTube and Android app download histories. Participants whose data had not exported were instructed to notify researchers when the data was ready. Communication

---

[5]All YouTube-related questions were asked for the last year.
[6]All options: "never," "rarely," "sometimes," "frequently," "always."

---

[7]takeout.google.com
[8]Mac users were asked to copy the content of the application folder and paste to a text-box in our upload portal. Windows users were asked to export relevant windows key registry entries with a bundle of PowerShell commands. Again they pasted the exported list to the portal text-box.

with participants was handled through the Prolific messaging system where no PII is needed.

Participants were finally informed about future procedures. Those who did/could not upload all data sources were instructed to notify us when data was ready for future tasks. Tutorial stage details are in the extended paper (Appendix A).

Data collection server & upload portal   In order to maintain state between stages and ensure minimal exposure of participants' data to third parties, we developed our data collection server and upload portal. Due to the relatively sensitive nature of the data collected, both pieces of software featured multiple security & privacy measures.

The upload portal operated on the client side and provided an interface to facilitate uploads to the data collection server (list of programs on computers, app download histories, and YouTube watch histories) and communicated what data sources we needed from participants (see the extended paper, Appendix A). The portal only retained files that were relevant to the study based on file names (history files and program lists). Retained data, along with a random 192-bit integrity nonce, was encrypted with researchers' public key before upload (we used an audited JavaScript port of NaCl [54][9]). The corresponding private key was only stored on two researchers' laptops. If participants attempted to upload a zip with relevant files, the portal parsed the zip and discarded any files we were not looking for (e.g., Apple exports contain data on physical addresses) before encryption and upload.

The data collection server (1) stored the uploaded encrypted data, (2) kept a tally of who signed up for the study (with anonymous prolific identifiers), and (3) stored participants' hardware configurations (Android, iOS, or "other" for mobile; Windows, Mac, or "other" for computers). All history and program data were received and stored in an encrypted state. We rate-limited the data collection server to prevent scans of the prolific ID space or attacks on our storage capabilities. We logged requests made to our endpoints along with associated prolific IDs for debugging and forensics purposes.

Researchers downloaded participants' data locally, decrypted it (with the private key stored only on two researchers' laptops), scraped additional data (video subtitles, engagement statistics, upload date, and genres) from YouTube.com and its API,[10] and analyzed it on their locally. The scraping tools are re-implementations of those described in [13].

3.1.3 Final stage

Once we received all data sources from participants or received confirmation (through Prolific messages) that exported data was ready to be provided, and responses/data passed our quality control tests (discussed below), participants were admitted to the final stage. This stage started with a prompt to upload any remaining data sources via the upload portal and continued with the final questionnaire. Participants who did not provide all required data sources were not allowed to continue. The rest were automatically directed to a questionnaire that contained a series of VPN-ad-related questions (Have you heard of VPNs before, where; Have you seen ads for VPNs before, where; and an open ended question on what they remember being advertised), our main mental models questions (the basis for construction of our main dependent variables, described in § 3.4.1), and demographics. We included two attention checks in the mental models questionnaires. Mental model questions were posed at the very end to limit bias in the earlier open-ended questions, where participants were asked to recall their knowledge. This arrangement is also likely to increase the quality of mental model measurements, as participants' thoughts are likely fresh in minds. A full list of questions is in the extended paper (Appendix A).

3.2 Piloting and preliminary study

In order to develop our mental models questionnaires, finalize our study design, and pick a reasonable modeling approach to the data we would collect, we conducted a series of pilots and preliminary studies.

Piloting   We piloted our study with three usable security researchers and three lay-users while developing procedures. These pilots helped us gauge the usability of our data collection and run end-to-end tests of the data collection infrastructure with various participant device configurations.

Mental models questionnaire development   We developed two questionnaires to measure mental models relevant to our research goals. After developing a set of statements for each questionnaire based on observations from [13], following methodology from prior researchers [55], we collected a series of responses from Prolific participants to (1) understand if participants were correctly interpreting our statements, (2) eliminate ceiling or floor effects, and (3) ensure that our statements were internally consistent. We asked open-ended questions after mental models questionnaires to gauge if participants were correctly interpreting our statements. After our iterations, we tweaked our Likert options, reworded some statements, and added attention checks. Our final scales did not exhibit ceiling or flooring effects and had acceptable internal consistency (Cronbach's α of .87, .84, .83, and .74) [56].

Preliminary study   After scale development and piloting, we collected a preliminary set of responses from 36 Prolific participants. Using this set, we explored different analysis techniques, finalized our definition of exposure to VPN ads, picked the covariates to consider in our analysis, and then pre-registered our exact data collection and analysis procedure.

Data quality control   Due to multiple dishonest and careless responses during our piloting and preliminary stages,

---

we implemented a series of quality control checks throughout our study. Participants had to give sensible responses to open-ended questions, pass attention checks, give quality data (YouTube histories consistent with survey responses (at least a year of data with no obvious gaps in the last year), non-empty list of installed programs, and non-empty app download histories). Participants who failed any of these checks were not included in the final dataset but often received partial compensation based on their time. Further, any participant we could not find influencer VPN ads (in YouTube histories) for but had self-reported to have seen 10 or more influencer VPN ads on YouTube in the last year were filtered out (n=3). We deemed that these participants had either given us missing data or were untruthful with survey responses. To incentivize completion, participants were paid for their time spent during the different stages of the study but only paid for the data they gave upon successful completion of the full study.

3.3 Ethical considerations

Our research objectives necessitate precise measurements of participants' exposure to VPN ads and various covariates that could influence mental models. We considered two alternatives when collecting this data: purely self-reported data, or observational data.

Watch histories For ad exposure, we preferred the latter. Dispite correlation with observational data self-reported ad exposure is highly error-prone [57, 58]. This margin of error would inhibit our nuanced analysis. Moreover, it is possible that ads may impact user mental models even if the users do not actively perceive it, let alone recall it years later. We collected participants' YouTube watch histories in order to measure their exposure to VPN ads. We asked participants to share histories from their personal accounts and screened out participants who shared their account with others.

App download histories and computer programs For a clearer link between mental models and VPN ad exposure, we tried to limit major confounds in our experimental setup. We suspected that previous use of VPNs might substantially influence mental models related to VPNs. In our preliminary study, self-reported VPN use resulted in less well fitting regression models than measured VPN use (from App installs and computer programs). This observation held true in the final data as well, including brand familiarity models. As with YouTube histories, participants were instructed to only share data from their personal devices—reducing the likelihood of infringing on non-consenting parties' privacy. For similar reasons, we did not ask for data from work devices. In hindsight, our analysis did not find owning a VPN to be statistically significant factor. Future work may chose to omit this variable.

Protective measures Due to the relatively sensitive nature of the data we collected, we took extra precautions to ensure participants were explicitly and implicitly aware of what information they were providing. We obtained formal informed consent for the screener and the full study. During the screener, after full procedures were described, participants were explicitly asked which data sources they were willing to share and whether they would like to continue with the study. We screened out participants who reported sharing their accounts with others and only asked about personal devices. When collecting data, we purposefully designed procedures such that participants had to follow tutorials that clearly stated what they were exporting and required them to manually give us the data, implying additional consent. Further, study stages were set up as independent tasks on Prolific, and each task was advertised with procedures and expected compensation, enabling reconsideration at any stage. One researcher actively monitored data collection and helped debug any issues that came up during the study. We compensated participants for extra time spent if we deemed their efforts to be made in good faith. To reduce exposure of participants' data to third parties (including system administrators), we developed our own data collection procedures and only kept fully decrypted watch/download history and computer program data on two researchers' laptops (see §3.1.2). Procedures were created iteratively, and all were approved by the University of Maryland IRB.

3.4 Measuring and defining variables.

Here we define the main independent and dependent variables used in our analysis, summarized in Table 2.

3.4.1 Mental models questionnaires

We measure participants' mental models by administering two Likert-type questionnaires: one for understanding of what VPNs are capable of, and a broader one for threats on the internet. Unlike measurements of exposure to VPN ads, using questionnaires to consistently measure mental models is still one of the only feasible solutions [26, 61]. Both scales were based on the most prominent observations from our previous work [13] but also ensured adequate coverage of the space of observed statements (e.g., we did not keep both "Governments are tracking you", and "Governments are tracking everyone," the two most popular themes in threats). The full set of statements along with what it measures are in Table 3.

VPN mental models Following observations from prior work, the first questionnaire attempted to measure a combination of prevalent statements about VPNs in VPN ads, as well as more specifically factual and misleading ones. These statements were all about VPNs directly, were presented together due to contextual relevance, and were asked using prevalence-based Likert-type options of "True for all VPNs," "True for almost all VPNs," "True for some VPNs," "True for almost no VPN," and "True for no VPN." We chose this framing

Variable	Explanation	Details
Dependent variables (DVs):		
Factual VPN mental models	(Dis)agreement with factual statements about VPNs featured in VPN ads.	§3.4.1
Misleading VPN mental models	(Dis)agreement with misleading statements about VPNs featured in VPN ads.	§3.4.1
All VPN mental models	(Dis)agreement with statements featured in VPN ads.	§3.4.1
Threat mental models	(Dis)agreement with threat statements featured in VPN ads. No mention of VPNs.	§3.4.1
VPN brand familiarity ($\times 5$)	Brand familiarity with ExpressVPN, NordVPN, ShurfShark, PIA, and Atlas VPN.	§3.4.3
Independent variables (IVs):		
Exposure to VPN ads	A measure of how much exposure to influencer VPN ads.	§3.4.2
VPN ownership	VPN software found in the list of programs or apps (didn't own).	§3.4.3
Technical expertise	How often participants are asked for tech advice. Bucketed into two (less often).	§3.4.3
Privacy sensitivity	Participants' IUIPC-8 scores.	[59, 60]
VPN ad interval	Average time between VPN ads.	§3.4.3

Table 2: Independent and dependent variables (IVs, DVs) used in analysis. Categorical variable baselines in parentheses.

as opposed to "Agree/Disagree" options because a large minority of participants interpreted agreement as prevalence during piloting and preliminary studies, while nearly all correctly interpreted the prevalence options. Scales were created out of Likert responses simply by adding up numeric values equivalent to the position of the selected Likert-type option (e.g., "True for all VPNs," → 5, "True for no VPN" → 1). The resulting scales for all mental models, misleading mental models, and factual mental models had Cronbach's alphas of .87, .84, and .74 respectively in a set of 88 preliminary Prolific responses. We deemed these to be acceptable, with results on factual mental models considered relatively tentative.

Threat mental models   The second questionnaire measured agreements with broader (hyperbolic) threat statements that were featured in VPN ads but were not specifically about VPNs [13]. We measured these threat statements separately from VPN-related ones since mental models of threats might have implications for security & privacy at a broader level. These statements are not necessarily tied to any specific technologies; thus, we asked participants' agreement using a seven point Likert from "Strongly agree," to "Strongly disagree." We created a six-item scale (see Table 3) and obtained a Cronbach's alpha of .83, once again, acceptable [56].

Although the threats we use in this scale contain some truth, we consider them hyperbolic: (1) ISPs track internet activities and sell derived data products; however, they cannot track *all* activity as most content is encrypted [62]. (2) Though it is hard to know exactly what the U.S. government does, the fourth amendment (and related statutes [63, 64]) does constrain surveillance to varying degrees [65]. (3) Due to end-to-end encrypted communications (most popular messaging apps) and data privacy measures, internet companies cannot collect *all* data. (4) To steal passwords and credit cards, hackers need to implement nuanced attacks, often requiring physical proximity (e.g., [66]), or sophisticated attacks on
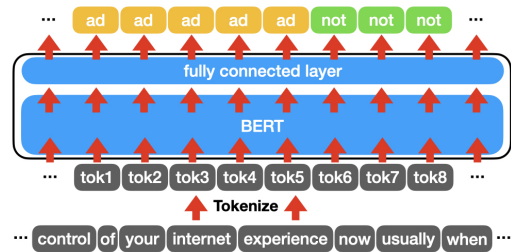

Figure 1: Ad classifier overview. In this example, "now usually when" is not part of an ad, the rest are.

identity/transaction management systems (e.g., [67]). This limits the effectiveness of such attacks, making them rare for the average user [68]. The threats mentioned are real problems, and we don't intend to minimize them here, but these scale items — and the ad contents they are drawn from — use language (e.g., all, everyone, easily) in ways that we consider hyberbolic.

3.4.2   Measuring exposure to ads

The second component we must be able to measure is our participants' exposure to influencer VPN ads. We base our exposure calculation on the total number of words that make up influencer VPN ads in the YouTube histories of our participants. We use the amount of words, as opposed to individual ad segment counts or aggregate duration of all influencer VPN ads, because we believe word count is a more accurate measure of information conveyed; some ads might be longer than others and some YouTubers might produce denser, faster-paced content compared to their peers.

To count VPN ad words, we augment BERT [69] with a fully connected output layer to produce an output vector of

Scale group	Staement
VPN mental models:	
Misleading	I don't have to worry about anything on the internet if I use a VPN.
Misleading	Companies can't collect my data on the internet if I use a VPN.
Misleading	My credit card information is protected online if I use a VPN.
Misleading	My passwords are protected online if I use a VPN.
Misleading	I am protected from seeing ads on the internet if I use a VPN.
Factual	My Internet Service Provider (e.g., Verizon, AT&T) can't find out which websites I go to if I use a VPN.
Factual	Hackers on the same wireless network as me can't see which websites I go to if I use a VPN.
Factual	I can watch other countries' streaming libraries (e.g., Netflix, Hulu) if I use a VPN.
Factual	I can overcome internet censorship if I use a VPN.
Factual	It seems like I'm browsing the internet from somewhere else, if I use a VPN.
Factual	A VPN encrypts my web traffic before it leaves my device.
Threat mental models:	
Threats	My Internet Service Provider (e.g., Xfinity, AT&T) is tracking all of my internet activity.
Threats	My Internet Service Provider (e.g., Xfinity, AT&T) is selling all of my internet activity.
Threats	The U.S. Government is tracking everyone (including me) online.
Threats	Internet companies are collecting all of my internet activity.
Threats	Hackers can easily steal my credit card information from me online.
Threats	Hackers can easily steal my passwords from me online.

Table 3: Likerts used for measuring mental models of threats and VPN capabilities. Item order was randomized for participants.

predictions given an input sequence of tokenized text (in our case, YouTube video transcripts). The objective is to output predictions that identify whether the corresponding input token (i.e., words or word parts) is part of a VPN ad or not. To parse videos that contain more than 512 tokens (a BERT limitation), we use a sliding window with a stride of 256, classifying tokens as part of an ad if either of the two predictions classify it as such. Once each word in a video is labeled, the problem reduces to a count of the ad-classified tokens. An overview of our approach is given in Figure 1.

To train our model, we use two datasets of video subtitles, each labeled per subtitle token as part of a VPN ad or not. The first dataset contains a random sample of 238 VPN ad and 238 non-VPN-ad videos (along with labels for where the ads are), from our prior work [13]. The second dataset was obtained by manually labeling 351 VPN ads from a crowd-sourced dataset of influencer YouTube ads[11]. Two coders labeled videos following the convention we previously set [13], identifying relevant videos with keyword searches, reaching agreement[12], and coding additional videos individually. We use the same dataset to include 351 non-VPN influencer ads in our dataset to ensure our classifier does not confuse such ads with VPN ads, a common pitfall in early experimentation. The overall dataset included labeled segments for both **short** (introductory or transient; e.g., "and a big thank you to nordvpn for sponsoring this video. If you liked [this] video make sure to

smash that like button") and **main** (about a minute long of describing threats online and benefits of VPNs) VPN segments as observed in our prior work [13].

Evaluation   We calculate several metrics for model assessment, including traditional metrics and a qualitative approach.

We first evaluate the model with five-fold cross-validation. On a per-word label basis, we achieve precision, recall, and f1 scores of .89, .88, and .88 respectively. Calculating the same metrics for each video in the validation set, we get an average precision, recall, and f1 of 0.92 ($\sigma = .20$), 0.96 ($\sigma = .11$), and .92 ($\sigma = .21$) respectively[13]. If we use any words' positive labeling within a video to equate to the video containing an ad, we achieve ad detection precision, recall, and f1 scores of 0.99, 0.94, and 0.97.

Next, to get a sense of how well this model would work on our participants' history dataset and to understand where it might not work as well, we take a semi-qualitative approach. We manually evaluate the model's predictions on a random sample of up to five videos per exploratory stage participant with at least one VPN-related phrase (e.g., "VPN," "Surf-Shark," "virtual private"), adding up to 138 unique videos. We call this set the *measurement validation set*. Two researchers independently searched through the videos and tried to identify VPN ads by (1) looking at common places where VPN ads appear, (2) keyword searches, and (3) the model inference results. They assigned one of two labels per ad segment to establish a ground truth set, 2 and 1, indicating the segment was

---

[11]https://sponsor.ajay.app/database

[12]We reached a Krippendorff's α of .98 over 300 ad segments, ($\alpha = .97$ over 165 videos), greater than the acceptable .85 from our previous work.

[13]As in prior work, we modify metrics to mitigate zero division errors [70].

in its entirety a VPN ad, or a VPN was part of the advertised product (henceforth, partial ads). We chose this qualitative way of labeling videos since the boundaries of VPN ads are often fuzzy and we wanted to differentiate the model's success on short segments compared to main segments. This style offered consistency with labels between researchers. The researchers additionally noted if the segments they labeled were main ad segments (usually a minute long, with the bulk of the content [13]) or short segments that were meant to remind viewers of the sponsor, or briefly mention the VPN product. They then repeated this process to judge all segments detected by the model. On participants' data, our model achieves precision, recall, and f1 score of .87, .90, and .88 respectively.[14]

Classifiers intended to detect security & privacy content are rare, making it hard to establish a fair baseline. However, we note that these numbers are arguably better than our previously reported agreement when we manually labeled videos for VPN ads (Krippendorff $\alpha = .85$, [13]), and on par with classifiers attempting to label privacy content with much larger language models (precision= .91, recall= .84 [71]).

Misclassifications We deem our models to be accurate enough for the type of analyses we plan to conduct. Regardless, misclassifications might introduce certain biases to our results. We notice three ad segment types that constitute nearly all misclassifications in the manual validation set: (1) short segments that usually do not contain much security & privacy-relevant content (16/32); (2) not being able to detect partial ad segments (6/32); and (3) segments that praise VPNs but are not explicitly sponsored (7/32). Notably, our metrics are significantly higher for the main ad segments, where most of the security & privacy relevant content is disseminated (recall= 0.98, precision= .90, f1= .94). Thus, most misclassifications (short segments, or partial ads) likely only impact our analysis of brand familiarity.

Defining exposure We define exposure to influencer VPN ads for participant i as:

$$\text{ad exposure}_i = \ln\left(\frac{\text{VPN ad word count}_i}{\frac{1}{N}\left(\sum_{k=1}^{N} \text{VPN ad word count}_k\right)} + 1\right)$$

where N denotes the number of videos (among all participants) that contain at least one ad-word. We use a log scale to account for the diminishing returns of repeated ads—the first watch will not necessarily have the same effect as the 50th. The utility of a logarithmic scale was observed in early ad response work [72] and has been adopted in various formulations of ad exposure since (e.g. [7, 11, 73]). Further, we had empirically observed (on preliminary data) that a log scale results in more accurate modeling on our exploratory data. The "+1" is to avoid undefined values for participants who do not have VPN ads in their histories.

3.4.3 Additional variables

Our preliminary analysis indicated that more than just exposure to ads might be correlated with mental models. These variables had either been found to potentially be correlated with mental models (p<.10) in our preliminary study or prior work. Specifically, we included the following covariates. We considered including demographic variables; however, aside from technical background, our preliminary study found no relationship between demographics and mental models.

Tech savviness We asked participants how often they gave tech advice, as a proxy for tech savviness [16, 61].

Privacy sensitivity Privacy sensitivity is likely to be associated with participants' understandings of VPNs and threats. We measure privacy sensitivity with IUIPC-8 [59, 60], a widely accepted privacy scale.

VPN ownership Our preliminary analysis showed VPN ownership history to correlate with mental models. We explored two measures of ownership: self-report and measured. Self-report data resulted in weaker-fitting models in the preliminary dataset; thus we used the measured approach (this is true for the final dataset, including brand familiarity models).

For completeness, we also explored including self-reported use of VPNs in work contexts in additional to personal use. This degraded the fit of the regression models in both the preliminary and final datasets, including brand familiarity models. Ultimately, we chose to focus on consumer VPNs (instead of corporate solutions) because (1) consumer VPNs were the focus of VPN ads [13], (2) corporate VPNs may conduct surveillance or other privacy-invasive features (e.g., [74]), (3) and corporate VPNs might be enabled by default (e.g., on *managed* devices [75–77]) without users realizing.

Since the most popular VPNs that advertised in our dataset are used via native apps,[15] we search the app download histories of participants' primary Android or Apple devices along with the list of programs installed on their primary computers, and consider a participant to have owned a VPN if we find a program that has "VPN," or any of 37 popular VPN brands (found in prior work, in online ranking lists) in the title.

Average VPN ad interval Prior work on traditional ads had found that the effectiveness of ads depends on how frequently customers are exposed to ads (e.g., [2, 78]). We measure the average days between viewing VPN ads for each participant using their YouTube watch histories.

Brand familiarity Prior work has consistently found a link between brand exposure and brand familiarity [2]. We explore the correlation between VPN ad exposure to specific brands and familiarity with that brand as a baseline check on the quality of our data. Using Likert-type scales ("not at all

---

[14]Due to our labeling style, regular VPN ad segments contribute twice as much to the scores (positively and negatively) as partial ads segments.

[15]Exact numbers are inaccessible but, these apps seemingly receive an order of magnitude more downloads than extension installs. E.g., 1M ExpressVPN extension downloads vs 50M+ ExpressVPN app installs

familiar" to "very familiar"), we measure familiarity with the five most popular brands from the preliminary study.

3.5 Analysis

Our goal is to investigate the relationship between ad exposure and mental models. As such, our main analysis consists of regressions on all four mental models variables (factual VPN, misleading VPN, all VPN, and threat mental models).

We do not try to model the exact shape of the relationship curve, as there is no clear consensus in the marketing literature on what the ad response function is (see §2, modeling ads). There seems to be consensus around wear-in/wear-out effects, and our preliminary analysis suggests the same. Quadratic terms have often been used to model this effect ([1–4]), but heavily more recently [79]. Thus, we simply test if the relationship between exposure to ads and mental models starts positive and turns negative after the maximum.

To achieve this, we specifically conduct *two lines* analysis proposed by Simonsohn [79] once per mental model variable. After finding the maxima, we fit two additional linear regression models per mental model variable to account for covariates (privacy sensitivity, tech savviness, VPN ownership, and average VPN ad interval): one before the maximum and one after. We do not include covariates in the two lines analysis since covariates might have different effects before/after the maxima, this in unaccounted for by two lines analysis.

We choose to include the following covariates in the set of independent variables: exposure to VPN ads, IUIPC-8, tech advice frequency (bucketed into two), ownership of VPNs, and mean time between exposure to ads. To avoid overfitting, we fit all possible models with combinations of these IVs that contain exposure to ads (our main variable of interest). We then select the adjusted-R^2 maximizing model for each DV. This analysis was repeated for each of the four mental models.

As a check on data quality, we largely use the same process as before but use brand familiarity as the dependent variable and exposure to the specific brand ad (same definition from before, but split by brands) as the main covariate. Unlike mental models, we do not run two lines analysis, we simply select (same selection strategy as before) one ordinal logistic regression per brand. We chose this approach since brand familiarity doesn't decrease with increased ad exposure [2]. Following prior work, the IVs were augmented with exposure to all other VPN brand ads [80].

We performed qualitative analysis on three short, open-ended questions in our survey: (1) what threats do VPNs protect against, (2) where did participants learn about VPNs, and (3) justifications for a response in the VPN mental models scale (selected at random for each participant). For each question, we established adequate inter-rater reliability metrics. All agreements were calculated over 10% of their respective datasets. For the first question (threats), two researchers used our existing codebook (obtained from [13]) of VPN and threat statements found in VPN advertisements. Following our prior approach [13], we coded the assets under threat, adversaries, and attacks per response to establish threat models. After obtaining a Krippendorff's α of .88 for combined codes (Strictly defined [13]. Individual subcode agreements are .94 for adversaries, .90 for attacks, and .97 for assets; all much higher than those previously reported [13]), researchers split the remaining responses evenly to code. Two researchers iteratively developed codebooks for the second (information source) and third (justifications) questions from scratch; after agreement was reached (Kupper-Hafner concordance of .92 [81, 82] and Krippendorff's α of .76 [56] respectively), one researcher coded all remaining responses.

To account for the sampling bias potentially introduced by the demanding nature of our study, we collect IUIPC-8 scores from (n=98) participants who ultimately did not finish the study. We compare the IUIPC-8 of this set and the final set of participants with a Mann-Whitney U test.

3.6 Limitations

We recruit our participants from a crowdsourced sample, limitations of which are well known: participants tend to be younger, more educated, and privacy-conscious than the general US population. Our measurements indicate that our participants might have been slightly less privacy-sensitive than general users on the platform, though our participants included a wide range of ages, educational backgrounds, and had a similar race distribution as the US public.

Our work focuses on influencer VPN ads, created by YouTubers and embedded directly in the video content (as opposed to interstitial YouTube ads). Participants could have received VPN ads through other means (e.g., TV ads, podcasts), or when they weren't logged into their personal accounts. This would confound our analysis. We screened out participants who often watched YouTube while not logged in, or who had others (e.g. partners) regularly watch YouTube on their accounts in their absence. Our results show that YouTube is the primary platform VPN ads appear by a large margin (see §4.1), indicating minimal outside platform influence.

Our questionnaires might not have distilled mental models accurately. We used the most prevalent statements from our representative sample of VPN ads [13], and extensively tested our questionnaires (multiple pilots and a preliminary study).

Our detection of VPN use is likely not perfect. Though the most popular brands in our dataset all had native applications, some also had extensions which our measurement wouldn't detect. The most popular VPNs in our dataset are primarily used through native apps, limiting this issue (see §3.4.3).

It is possible that some VPN ads were too nuanced for our classifier to detect. Based on our assessments (§3.4.2), we believe our model is sufficient for our analysis.

We did not and could not measure exposure to VPN ads outside of YouTube. Similarly, we limit causal arguments as

Gender	Female	98
	Male	105
	Self described	14
Age	18-25	53
	26-35	82
	36-45	47
	46-60	25
	61+	10
Ethnicity	White	141
	Black or African Am.	17
	Asian or Asian Am.	23
	Hispanic or Latino	12
	Other or mixed race	24
Education	Completed H.S. or below	28
	Some college, no degree	38
	Trade or vocational	5
	Associate's degree	26
	Bachelor's degree	97
	Master's or higher degree	23
YouTube videos in history	<10K	50
	10K-30K	61
	30K-50K	92
	50K+	14
Give technology advice	Always, often	73
	Sometimes, rarely, never	144
VPN ownership	Yes	77
	No	140

Table 4: Participant demographics.

we cannot measure every variable that might explain VPN and security & privacy mental models. However, our purpose-built questionnaires captured the most common claims in VPN ads, and controlled for major confounding variables.

Our modeling of exposure is likely incomplete. On the preliminary dataset we experimented with several measures from prior work, ours produced the best models. Our goal isn't necessarily to find the best model but a useful one.

US participants were recruited for this study. Our results might not generalize well to populations with different cultures and societal norms. This is an important limitation, as a large portion of the VPN market is outside of the US [83].

4 Results

We present our results in this section. First, we characterize our participants. Then, we present our main analyses.

4.1 Demographics/participants

In total, we screened 2755 participants, ∼830 of which met our eligibility criteria and were invited. Ultimately, 217 participants completed our full study in August/September 2023.

While we collected data across a diverse pool of ages and education levels, our participants skewed younger and more educated compared to the US population (see Table 4).

The privacy sensitivities (IUIPC-8) of participants who finished the study were significantly lower than those who wished to not continue (p=0.03) with a location shift estimate of 2.00 (IUIPC-8 range: [7, 56]). In contrast, no significant difference was found (p=0.07, location shift estimate: 1.00) between participants who finished the study and those who were screened out (wished to not continue, incompatible hardware, low YouTube usage). This plausible selection bias might mitigate the much higher privacy sensitivities of Prolific participants compared to the general population [84], [85].

Watching YouTube and VPN ads   The mean length of our participants' YouTube histories was 6.42 years (min=1.01, max=13.07, σ=3.6). On average, participants watched 27.00K (min=0.43K, max=90.31K, σ=16.98K) YouTube videos with 82.26 (min=0, max=1050, σ=124.42) containing VPN ads.

Self-reported responses align with these measurements of YouTube use and VPN ad exposure. When asked to recall (open-ended) where they heard about VPNs, 52.5% of participants volunteered ads, 37.7% mentioned YouTube, and 29.5% mentioned ads/sponsorships on YouTube. Among participants who said they had seen VPN ads embedded in videos (165 out of 217), 160 (97.0%) selected YouTube as the medium, implying that YouTube is the primary distributor such ads. In contrast, 12.7% selected Twitter, 10.9% TikTok, 10.9% Instagram, 8.4% Facebook, and 12.7% selected TV.

4.2 Brand familiarity and VPN ad exposure

To start to understand our data, we explore the relationship between exposure to VPN ads and familiarity with VPN brands via ordinal logistic regressions. As described in §3.5 we select the best fitting brand familiarity model for each of the top five VPNs that appeared in the preliminary data. We report confidence intervals for each coefficient and, due to its similarity with OLS R^2, we report Nagelkerke's pseudo-R^2 [86], [87].

The results show that every order of magnitude increase in exposure (defined on a log scale) to ExpressVPN, NordVPN, Surfshark, Private Internet Access (PIA), and Atlas VPN ads leads to an increased chance of familiarity with that specific brand by a factor of 1.41, 1.51, 2.30, 1.98, and 1.70, respectively. Selected models for each brand are shown in Table 5; brand exposure variables for each model were significant predictors of familiarity (visible in Figure 2).

Being less tech-savvy is significantly associated with lower familiarity with NordVPN, and was selected in SurfShark and PIA models. Higher privacy sensitivity is significantly associated with higher familiarity with NordVPN, and appears in the ExpressVPN model. This likely indicates that privacy sensitive or tech savvy people are more familiar with security & privacy products. No other covariate is significant

	ExpressVPN		NordVPN		SurfShark		PIA		AtlasVPN	
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Brand exp.	1.41*	[1.01, 1.96]	1.51***	[1.27, 1.79]	2.30***	[1.68, 3.19]	1.98***	[1.38, 2.86]	1.7**	[1.23, 2.35]
IUIPC-8 score	1.03	[0.99, 1.08]	1.05*	[1.01, 1.10]						
Low tech adv.			0.45**	[0.27, 0.76]	0.98	[0.54, 1.80]	0.91	[0.47, 1.81]		
Had VPN			1.37	[0.82, 2.27]					0.94	[0.48, 1.80]
Avg. ad interval	1.00	[1.00, 1.00]								
Non-brand exp.	1.19	[0.82, 1.75]			1.07	[0.88, 1.34]	0.96	[0.85, 1.11]		
Ad count	7631		3784		2558		633		442	
AIC	702.8		795.1		538.6		437.3		397.2	
R2 Nagelkerke	0.42		0.21		0.31		0.08		0.06	

Table 5: Ordinal logistic regression models predicting familiarity with VPN brands. OR: odds ratios, exp.: VPN ad exposure.

	Threats		Factual		Misleading		All VPN statements	
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
VPN ad exposure	0.84***	[0.35, 1.33]	0.06	[-0.24, 0.37]	-0.41+	[-0.87, 0.05]	-0.22	[-0.83, 0.40]
IUIPC-8 scores	0.12+	[-0.01, 0.25]	0.07+	[-0.01, 0.15]			0.07	[-0.09, 0.24]
Low tech adv.	-1.38	[-3.15, 0.38]	-0.87	[-1.97, 0.23]				
Had VPN	-1.01	[-2.75, 0.73]	0.99+	[-0.10, 2.08]	-0.80	[-1.99, 0.40]		
Avg. VPN ad exposure					0.00	[-0.01, 0.00]		
AIC	1402.5		1200.2		1120.6		1515.2	
R^2	0.089		0.049		0.026		0.005	

Table 6: Linear regression models for mental models. +: $p < 0.1$, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$

or consistent in its effect direction between models, preventing further takeaways. Nonetheless, some were selected in various individual models: ad interval was selected in the ExpressVPN model, and "having a VPN installed" appeared in the NordVPN and AtlasVPN models.

To the best of our knowledge, this is the first time security & privacy awareness has been directly linked to security & privacy media exposure outside of self-report measures. Our dataset likely captured the fundamental desired effect of ads: establishing familiarity with a brand.

4.3 Mental models and exposure to VPN ads

After we establish a baseline relationship between exposure to VPN ads and what participants think, we move on to our main analysis. Following our plan, we first conduct two lines analysis for each of our four mental model variables with respect to exposure to VPN ads. This analysis does not show clear effects, the potential wear-out effects do not seem to lead to a significant ($p>0.05$) decline in mental model metrics after a breakpoint. Conversely, the estimated regression lines for threat and misleading VPN mental models (Figure 3) suggests a weak but consistent trend across the VPN ad exposure range. Thus, we modify our analysis plan: we fit one (instead of two) linear regression model for the entire VPN ad exposure range for each mental model variable, similar to the analysis conducted with brand familiarity. We follow the same model selection procedure as described before (§3.5).

Our final results are given in Table 6. The selected model for threat mental models shows a significant and strong relationship between exposure to VPN ads and threat mental models: leading to an estimated 5.7 point change in threat mental models (in an effective range of 35) between the participants with the least and most exposure. This effect exists even though the selected model controls for privacy sensitivity (IUIPC-8), technical expertise (tech advice frequency), and VPN ownership history.

No other variable was a significant predictor of any mental model variable; however, we note the following observations from selected covariates: higher factual VPN mental models might be associated with higher privacy sensitivity (p=0.08), technical expertise (p=0.12), and owning a VPN (p=0.07); higher misleading VPN mental models might be associated with lower VPN ad exposure (p=0.08) and not owning a VPN (p=0.19). Mean VPN ad interval was selected in the Misleading mental models model and IUIPC-8 was selected in all VPN mental models model but neither was significant.

To understand users' mental models of VPNs in greater detail, each participant was asked to justify one answer selected at random from the Likert-type questions about VPN mental models (see Table 3). This qualitative analysis supplements our quantitative analysis by highlighting aspects of user's VPN mental models that closed-ended questions might not capture. Two researchers qualitatively coded these 217 free responses into one of five categories that captured how users justified their Likert-type answer: some VPNs offer
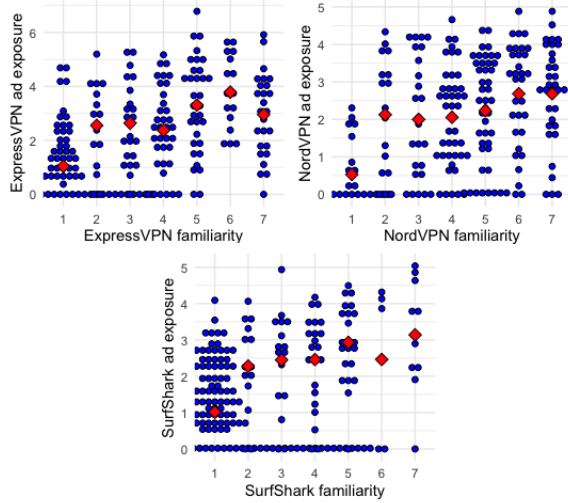
Figure 2: Brand familiarity and VPN ad exposure for each brand. Red diamonds denote the means. Familiarity ranges from "not at all familiar" (1) to "very familiar" (7).
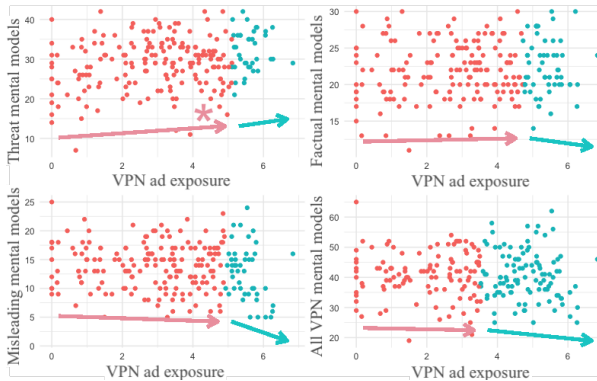

Figure 3: Two-lines analysis. Regression lines (separated by the breakpoint) were color-coded and shifted for clarity. Left regression on threat mental models fits significantly.

different features than others (20.7%), VPNs improve things in the average case (13.4%), VPNs are incapable of performing a task (25.8%), the task is an inherent function of a VPN (32.3%), or participant's free response conflicted with or failed to justify their Likert-type response (7.8%). 12.4% (10/81) of participants who were asked about misleading statements said that was an inherent feature of VPNs, though most of these respondents expressed having a lack of knowledge on the subject and erred towards VPNs being capable of accomplishing tasks. 15.6% (10/32) of respondents who were asked about features often bundled with VPNs (such as password leak notifications) thought that those features were not add-ons, but core VPN capabilities. The opposite held true as well; 9.6% (10/104) of respondents who were asked about factual claims *underestimated* the capabilities of VPNs. These included mis-
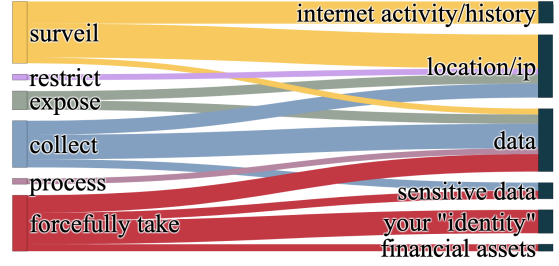

Figure 4: Attack-asset pairs participants said VPNs would protect against. Band width corresponds to number of responses. Pairs with fewer than five responses not shown.

conceptions of encryption, such as "a VPN does not change or encrypt traffic, only reroutes it" and "I believe it encrypts at their data centers, not my device." Finally, 10 participants explicitly cited advertisements as a source of knowledge, suggesting VPN ad impact, albeit small, on mental models.

4.4 What do VPNs protect against?

To better understand threat protection perceptions, we asked participants to list (open-ended) the two most severe threats they believe VPNs protect against. This question allows us to compare threat models of our participants to those advertised in VPN ads [13]. We analyzed responses using the same methods we outlined previously [13], determining attacks, adversaries, and assets. We find that while there is significant overlap in all three categories, notable differences do exist.

Adversaries Similar to VPN ads, participants most frequently named "hackers" (8.2% of participants), vague adversaries (e.g., "bad guys", 8.0% of participants), and vague companies (e.g., "companies", 4.0%) as the adversaries VPNs protect against. Surprisingly, our participants gave much more emphasis on malware compared to VPN ads (5.2% of participants vs. none noted [13]). However, unlike VPN ads, participants rarely named governments (1.7%) or ISPs (2.4%) as adversaries. Further, 66.8% did not name any adversary. The focus on malware might be explained by end users commonly using antivirus solutions [88].

Attacks and assets In contrast to adversaries, participants were much more likely to list assets protected (64.7%) and threats mitigated by VPNs (72.1%), often in the same response. Figure 4 depicts this relationship.

Participants listed many fewer unique assets, with different frequency, than ads did. Location/IP protection was most commonly noted (18.5%), often associated with surveillance, unwanted exposure and collection. "Data" (14.7%) and "sensitive data" (5.8%) was often collected, or forcefully taken (e.g., "stolen"). Many participants were convinced that VPNs would protect their identities (5.6%), especially against identity theft (4.6%). In contrast, our prior work reports that internet activity is the most commonly mentioned asset in VPN ads,

followed by "data", nebulous security & privacy concepts (e.g., "privacy," "safety"), and "yourself." Identity theft does not appear to be a common threat in VPN ads.

The (non)impact of VPN ad exposure Though we found similarities between the threat models of our participants and those advertised in VPN ads, we hypothesize that this effect should be stronger among those who were exposed to more VPN ads. To test this, we first obtained the popularity ranking of each asset, attack, and adversary among VPN ads from our previous work [13]. We then ran a series of regression analyses between exposure to VPN ads and the ranking of assets, attacks, and adversaries reported by our participants within the VPN ad popularity list. We expected statements from participants with more exposure to rank higher. However, we find no statistically significant relationship, implying that even if this effect exists, it is likely weak.

5 Concluding discussion

We explored the impact of VPN ads on users' mental models. We measured mental models through specially developed questionnaires and measured exposure to ads by analyzing users' YouTube histories with a purpose-built BERT-based ad detector. Our results show that YouTube users are extensively exposed to VPN ads, which we found to strongly correlate with brand familiarity and increased belief in hyperbolic threats advertised. However, we find no significant relation between ad exposure and specific mental models of VPNs, this includes misleading mental models (Table 3). We discuss the implications of our findings below.

Security & privacy media exposure correlates with mental models: a new form of evidence Our analysis suggests that VPN ads likely impact users, partially confirming our prior speculation [13]. Participants who saw more ads for any of the five specific VPN brands also were significantly more likely to be more familiar with the respective brand. Further, we find that increased belief in (hyperbolic) threat statements made in VPN ads is linked with increased exposure to VPN ads. To the best of our knowledge, this is the first time exposure to security & privacy media was measured through non-self-report means and explicitly linked to mental models, confirming findings from previously exclusively self-report studies [15–17, 21].

Correlation with threats but not VPN mental models Unlike threat mental models, our analysis does not provide conclusive evidence linking exposure to VPN ads with increased belief in specific VPN mental models expressed in VPN ads. This observation may be due to emotional appeals (such as those in threat statements [13]) in ads having a stronger effect on consumers than technical appeals [22]. Our qualitative results suggest that the technical information in VPN ads might affect users (they answered VPN-related questions while referencing ads as their source), but isn't potent enough to produce statistically significant results.

What do VPN ads do We argue that the relationship between brand familiarity and exposure is due to ads increasing familiarity with a brand. This phenomenon is well studied in prior work [2]. This relationship isn't as straightforward with mental models. Ads could be contributing to the development of these models, or reinforcing them for users who already hold them. Regardless, our results show that while technical details might not be memorable, threat models are, perhaps as a result of seeing ads or perhaps because advertisers (or YouTube content delivery algorithms) target users with these models. Awareness of threats can benefit users via increased vigilance, but hyperbole can create excessive fear; further research is needed to determine what the right balance is.

Misleading mental models We observe that a large number of participants (see Figure 3, bottom left) believe in misleading mental models about VPNs. This observation has been echoed in prior work [26] Though our work does not find a direct correlation between these the misleading mental models and exposure to VPN ads that mention these models [13], this does not mean they are harmless. These ads might still mislead mental models, but in less obvious ways. For instance, users might increase confidence in their misinformed mental models through exposure to such ads.

Misinformed mental models may result in poor decision making when adopting VPNs. For instance, users might purchase a VPN when they don't need one (e.g., using TOR over VPNs [28]); or worse, might think they are protected against certain threats when they aren't (e.g., thinking VPNs prevent credit card misuse [26]).

Consumer education Misinformed mental models could perhaps be mitigated through consumer education. Our work, along with prior work [89], hints that technical details might not be memorable in short interventions. Further, it indicates that interstitial media might not be the right platform to communicate nuanced security mental models. Similar to the public health approach to medicine, early intervention before undesirable outcomes occur might be necessary (e.g., [90, 91]). As previous research has noted [61, 92], perhaps early educational curricula should incorporate appropriate use cases for security & privacy tools.

Recommendations for future work Though our study provides concrete evidence of the relationship between security & privacy media and mental models, it is not without limitations. We do not explore a multitude of variables that could affect advertising effectiveness [2]. Further, our study is limited to YouTube and VPN ads. Though we expect similar results from other security & privacy media, this is not a foregone conclusion. By exploring these factors, researchers can clarify the exact limits of security & privacy media influence.

Acknowledgments

References

[1] S. Y. Lee and Y.-S. Cho, "Exploring wearin and wearout in web advertising: the role of repetition and brand familiarity," *International Journal of Electronic Marketing and Retailing*, 2010.

[2] S. Schmidt and M. Eisend, "Advertising repetition: A meta-analysis on effective frequency in advertising," *Journal of Advertising*, 2015.

[3] P. Chatterjee, D. L. Hoffman, and T. P. Novak, "Modeling the clickstream: Implications for web-based advertising efforts," *Marketing Science*, 2003.

[4] D. Cox and A. D. Cox, "Beyond first impressions: The effects of repeated exposure on consumer liking of visually complex and simple product designs," *Journal of the Academy of Marketing Science*, 2002.

[5] J. Deighton, C. M. Henderson, and S. A. Neslin, "The effects of advertising on brand switching and repeat purchasing," *Journal of marketing research*, 1994.

[6] A. Hassan and S. J. Barber, "The effects of repetition frequency on the illusory truth effect," *Cognitive Research: Principles and Implications*, 2021.

[7] P. Manchanda, J.-P. Dubé, K. Y. Goh, and P. K. Chintagunta, "The effect of banner advertising on internet purchasing," *Journal of Marketing Research*, 2006.

[8] H. Simon, "Adpuls: An advertising model with wearout and pulsation," *Journal of Marketing Research*, 1982.

[9] K. R. Betts, K. J. Aikin, B. J. Kelly, M. Johnson, S. Parvanta, B. G. Southwell, N. Mack, J. Tzeng, and L. Cameron, "Taking repeated exposure into account: An experimental study of direct-to-consumer prescription drug television ad effects," *Journal of Health Communication*, 2019.

[10] I. Yaveroglu and N. Donthu, "Advertising repetition and placement issues in on-line environments," *Journal of Advertising*, 2008.

[11] I. Chae, H. A. Bruno, and F. M. Feinberg, "Wearout or weariness? measuring potential negative consequences of online ad volume and placement on website visits," *Journal of Marketing Research*, 2019.

[12] J. T. DeFrank, N. D. Berkman, L. Kahwati, K. Cullen, K. J. Aikin, and H. W. Sullivan, "Direct-to-consumer advertising of prescription drugs and the patient–prescriber encounter: a systematic review," *Health communication*, 2020.

[13] O. Akgul, R. Roberts, M. Namara, D. Levin, and M. L. Mazurek, "Investigating influencer vpn ads on youtube," in *IEEE Symposium on Security and Privacy (SP)*, 2022.

[14] DJ Cook, "Meet the Fortnite Scammers," YouTube, Feb 2020, https://www.youtube.com/watch?v=PWg_-Fjja3Y&t=91s.

[15] K. R. Fulton, R. Gelles, A. McKay, Y. Abdi, R. Roberts, and M. L. Mazurek, "The effect of entertainment media on mental models of computer security," in *Proc. SOUPS*. USENIX Association, 2019.

[16] E. M. Redmiles, S. Kross, and M. L. Mazurek, "How i learned to be secure: A census-representative survey of security advice sources and behavior," in *Proc. CCS*, 2016.

[17] K. Baig, E. Kazan, K. Hundlani, S. Maqsood, and S. Chiasson, "Replication: Effects of media on the mental models of technical users," in *Proc. SOUPS*. USENIX Association, 2021.

[18] M. Tahaei, A. Jenkins, K. Vaniea, and M. Wolters, ""i don't know too much about it": On the security mindsets of computer science students," in *STAST 2019*. Springer, 2021.

[19] E. M. Redmiles, A. R. Malone, and M. L. Mazurek, "I think they're trying to tell me something: Advice sources and selection for digital security," in *IEEE Symposium on Security and Privacy (SP)*, 2016.

[20] M. Namara, D. Wilkinson, K. Caine, and B. P. Knijnenburg, "Emotional and practical considerations towards the adoption and abandonment of vpns as a privacy-enhancing technology," *Proc. on PETs*, 2020.

[21] S. Ruoti, T. Monson, J. Wu, D. Zappala, and K. Seamons, "Weighing context and trade-offs: How suburban adults selected their online security posture," in *Proc. SOUPS*. USENIX Association, 2017.

[22] C. Obermiller, E. Spangenberg, and D. L. MacLachlan, "Ad skepticism: The consequences of disbelief," *Journal of advertising*, 2005.

[23] N. A. Jones, H. Ross, T. Lynam, P. Perez, and A. Leitch, "Mental models: an interdisciplinary synthesis of theory and methods," *Ecology and society*, 2011.

[24] R. Ramesh, A. Vyas, and R. Ensafi, "All of them claim to be the best": Multi-perspective study of vpn users and vpn providers," in *Proc. USENIX Security*. USENIX Association, 2023.

[25] A. Dutkowska-Zuk, A. Hounsel, A. Morrill, A. Xiong, M. Chetty, and N. Feamster, "How and why people use virtual private networks," in *Proc. USENIX Security*. USENIX Association, 2022.

[26] P. Story, D. Smullen, Y. Yao, A. Acquisti, L. F. Cranor, N. Sadeh, and F. Schaub, "Awareness, adoption, and misconceptions of web privacy tools," *Proceedings on Privacy Enhancing Technologies*, 2021.

[27] Y. Zou, K. Roundy, A. Tamersoy, S. Shintre, J. Roturier, and F. Schaub, "Examining the adoption and abandonment of security, privacy, and identity theft protection practices," in *Proc. of ACM CHI*, 2020.

[28] M. Fassl, A. Ponticello, A. Dabrowski, and K. Krombholz, "Investigating security folklore: A case study on the tor over vpn phenomenon," *Proc. ACM Human-Computer Interaction*, no. CSCW2, 2023.

[29] M. Ikram, N. Vallina-Rodriguez, S. Seneviratne, M. A. Kaafar, and V. Paxson, "An analysis of the privacy and security risks of android vpn permission-enabled apps," in *Proc. IMC*, 2016.

[30] C. Farivar, "FTC must scrutinize hotspot shield over alleged traffic interception, group says," *Ars Technica*, 2017, https://arstechnica.com/tech-policy/2017/08/ftc-must-scrutinize-hotspot-shield-over-alleged-traffic-interception-group-says/.

[31] R. Ramesh, L. Evdokimov, D. Xue, and R. Ensafi, "VP-Nalyzer: Systematic Investigation of the VPN Ecosystem," in *Network and Distributed System Security*. The Internet Society, 2022.

[32] Z. Weinberg, S. Cho, N. Christin, V. Sekar, and P. Gill, "How to catch when proxies lie: Verifying the physical locations of network proxies with active geolocation," in *Proc. of IMC*. Association for Computing Machinery, 2018.

[33] P. Bischoff, ""zero logs" VPN exposes millions of logs including user passwords, claims data is anonymous," *Comparitech*, 2020, https://www.comparitech.com/blog/vpn-privacy/ufo-vpn-data-exposure/.

[34] A. Ng, "How private is my vpn?" *The Markup*, 2021, https://themarkup.org/ask-the-markup/2021/08/12/how-private-is-my-vpn.

[35] M. T. Khan, J. DeBlasio, G. M. Voelker, A. C. Snoeren, C. Kanich, and N. Vallina-Rodriguez, "An empirical analysis of the commercial VPN ecosystem," in *Proc. IMC*, 2018.

[36] S. Kross, E. Hargittai, and E. M. Redmiles, "Characterizing the online learning landscape: What and how people learn online," *Proc. ACM Human-Computer Interaction*, no. CSCW1, 2021.

[37] A. Smith, S. Toor, and P. Van Kessel, "Many turn to youtube for children's content, news, how-to lessons," *Pew Research Center*, 2018, https://www.pewresearch.org/internet/2018/11/07/many-turn-to-youtube-for-childrens-content-news-how-to-lessons/.

[38] C. Lou, S.-S. Tan, and X. Chen, "Investigating consumer engagement with influencer-vs. brand-promoted ads: The roles of source and disclosure," *Journal of Interactive Advertising*, 2019.

[39] S. Nazerali, "How YouTube influencers are rewriting the marketing rulebook," *Huffington Post*, 2017, https://www.huffpost.com/entry/how-youtube-influencers-are-rewriting-the-marketing_b_59d2b250e4b03905538d17c3.

[40] S. Yuan and C. Lou, "How social media influencers foster relationships with followers: The roles of source credibility and fairness in parasocial relationship and product interest," *Journal of Interactive Advertising*, 2020.

[41] C. Lou and H. K. Kim, "Fancying the new rich and famous? explicating the roles of influencer content, credibility, and parental mediation in adolescents' parasocial relationship, materialism, and purchase intentions," *Frontiers in psychology*, 2019.

[42] M. Ali, A. Goetzen, A. Mislove, E. M. Redmiles, and P. Sapiezynski, "Problematic advertising and its disparate exposure on facebook," in *Proc. USENIX Security*. USENIX Association, 2023.

[43] M. Ali, P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, and A. Rieke, "Discrimination through optimization: How facebook's ad delivery can lead to biased outcomes," *Proc. ACM Human-Computer Interaction*, no. CSCW, 2019.

[44] E. Zeng, M. Wei, T. Gregersen, T. Kohno, and F. Roesner, "Polls, clickbait, and commemorative $2 bills: problematic political advertising on news and media websites around the 2020 u.s. elections," in *Proc. of IMC*. Association for Computing Machinery, 2021.

[45] Z. Moti, A. Senol, H. Bostani, F. Z. Borgesius, V. Moonsamy, A. Mathur, and G. Acar, "Targeted and troublesome: Tracking and advertising on children's websites," in *IEEE Symposium on Security and Privacy (SP)*, 2024.

[46] L. F. Cranor, "A framework for reasoning about the human in the loop," in *Proc. UPSEC*, 2008.

[47] C. Bravo-Lillo, L. Cranor, S. Komanduri, S. Schechter, and M. Sleeper, "Harder to Ignore? Revisiting Pop-Up Fatigue and Approaches to Prevent It," in *Proc. SOUPS*, 2014.

[48] Y. Albayram, M. M. H. Khan, T. Jensen, and N. Nguyen, ""...better to use a lock screen than to worry about saving a few seconds of time": Effect of Fear Appeal in the Context of Smartphone Locking Behavior," in *Proc. SOUPS*. USENIX Association, 2017.

[49] E. A. Qahtani, M. Shehab, and A. Aljohani, "The Effectiveness of Fear Appeals in Increasing Smartphone Locking Behavior among Saudi Arabians," in *Proc. SOUPS*. USENIX Association, 2018.

[50] S. R. Gopavaram, J. Dev, S. Das, and J. Camp, "Iotmarketplace: Informing Purchase Decisions with Risk Communication," Indiana University Bloomington, Tech. Rep., 2019.

[51] P. Emami-Naeini, J. Dheenadhayalan, Y. Agarwal, and L. F. Cranor, "Are consumers willing to pay for security and privacy of iot devices?" in *Proc. USENIX Security*. USENIX Association, 2023.

[52] C. Pechmann and D. W. Stewart, "Advertising repetition: A critical review of wearin and wearout," *Current issues and research in advertising*, 1988.

[53] K. Lehnert, B. D. Till, and B. D. Carlson, "Advertising creativity and repetition: Recall, wearout and wearin effects," *International Journal of Advertising*, 2013.

[54] D. J. Bernstein, "Cryptography in nacl," *Networking and Cryptography library*, 2009.

[55] D. Votipka, D. Abrokwa, and M. L. Mazurek, "Building and validating a scale for secure software development self-efficacy," in *Proc. of ACM CHI*, 2020.

[56] K. Krippendorff, "Reliability in content analysis: Some common misconceptions and recommendations," *Human communication research*, 2004.

[57] L. Vavreck *et al.*, "The exaggerated effects of advertising on turnout: The dangers of self-reports," *Quarterly Journal of Political Science*, 2007.

[58] A. R. Romberg, M. Bennett, S. Tulsiani, B. Simard, J. M. Kreslake, D. Favatas, D. M. Vallone, and E. C. Hair, "Validating self-reported ad recall as a measure of exposure to digital advertising: an exploratory analysis using ad tracking methodology," *International Journal of Environmental Research and Public Health*, 2020.

[59] 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*, 2004.

[60] T. Groß, "Validity and reliability of the scale internet users' information privacy concerns (iuipc)," *Proceedings on Privacy Enhancing Technologies*, 2021.

[61] O. Akgul, W. Bai, S. Das, and M. L. Mazurek, "Evaluating in-workflow messages for improving mental models of end-to-end encryption," in *Proc. USENIX Security*. USENIX Association, 2021.

[62] Google, "HTTPS encryption on the web," 2024, accessed: 2024-06-11. https://transparencyreport.google.com/https/overview?hl=en.

[63] U.S. Constitution amendment IV.

[64] Foreign Intelligence Surveillance Act, 1978.

[65] J. Kelley, "Privacy isn't dead – far from it," 2024, https://www.eff.org/deeplinks/2024/02/privacy-isnt-dead-far-it.

[66] M. Vanhoef and F. Piessens, "Key reinstallation attacks: Forcing nonce reuse in wpa2," in *Proc. of CCS*, 2017.

[67] Majority Staff for Chairman Rockefeller, "A "kill chain" analysis of the 2013 target data breach," United States Senate Committee on Commerce, Science, and Transportation, Tech. Rep., 2014, https://www.commerce.senate.gov/services/files/24d3c229-4f2f-405d-b8db-a3a67f183883.

[68] C. Breen, C. Herley, and E. M. Redmiles, "A large-scale measurement of cybercrime against individuals," in *Proc. of ACM CHI*, 2022.

[69] J. Devlin, M.-W. Chang, K. Lee, and K. N. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[70] R. Usbeck, M. Röder, A.-C. Ngonga Ngomo, C. Baron, A. Both, M. Brümmer, D. Ceccarelli, M. Cornolti, D. Cherix, B. Eickmann *et al.*, "Gerbil: general entity annotator benchmarking framework," in *Proc. of WWW*, 2015.

[71] O. Akgul, S. T. Peddinti, N. Taft, M. L. Mazurek, H. Harkous, A. Srivastava, and B. Seguin, "A decade of privacy-relevant android app reviews: Large scale trends," *arXiv preprint arXiv:2403.02292*, 2024.

[72] V. R. Rao, "Alternative econometric models of sales-advertising relationships," *Journal of Marketing Research*, 1972.

[73] P. J. Danaher and T. S. Dagger, "Comparing the relative effectiveness of advertising channels: A case study of a multimedia blitz campaign," *Journal of Marketing Research*, 2013.

[74] Cisco, "What is employee monitoring?" accessed: 2024-06-11. https://www.cisco.com/c/en/us/solutions/hybrid-work/what-is-employee-monitoring.html.

[75] Apple Inc., "Managing devices and corporate data," 2022, accessed: 2024-06-11. https://www.apple.com/business/docs/resources/Managing_Devices_and_Corporate_Data.pdf.

[76] ——, "Vpn overview for apple device deployment," 2024, accessed: 2024-06-11. https://support.apple.com/guide/deployment/vpn-overview-depae3d361d0/web.

[77] Microsoft, "Windows 10/11 and windows holographic device settings to add vpn connections using intune," 2024, accessed: 2024-06-11. https://learn.microsoft.com/en-us/mem/intune/configuration/vpn-settings-windows-10.

[78] P. Malaviya and B. Sternthal, "The persuasive impact of message spacing," *Journal of Consumer Psychology*, 1997.

[79] U. Simonsohn, "Two lines: A valid alternative to the invalid testing of u-shaped relationships with quadratic regressions," *Advances in Methods and Practices in Psychological Science*, 2018.

[80] B. Libai, E. Muller, and R. Peres, "The role of within-brand and cross-brand communications in competitive growth," *Journal of Marketing*, 2009.

[81] N. Malkin, A. F. Luo, J. Poveda, and M. L. Mazurek, "Optimistic Access Control for the Smart Home," in *IEEE Symposium on Security and Privacy (SP)*, 2023.

[82] M. Harbach, A. De Luca, N. Malkin, and S. Egelman, "Keep on Lockin' in the Free World: A Multi-National Comparison of Smartphone Locking," in *Proc. of ACM CHI*, 2016.

[83] atlasVPN, "Global vpn adoption index," 2020, accessed 2023-12-03. https://atlasvpn.com/vpn-adoption-index.

[84] J. Tang, E. Birrell, and A. Lerner, "Replication: How well do my results generalize now? the external validity of online privacy and security surveys," in *Proc. SOUPS*. USENIX Association, 2022.

[85] D. Abrokwa, S. Das, O. Akgul, and M. L. Mazurek, "Comparing security and privacy attitudes between ios and android users in the us," in *Proc. SOUPS*. USENIX Association, 2021.

[86] N. J. Nagelkerke *et al.*, "A note on a general definition of the coefficient of determination," *biometrika*, 1991.

[87] D. A. Walker and T. J. Smith, "Nine pseudo r2 indices for binary logistic regression models," *Journal of Modern Applied Statistical Methods*, 2016.

[88] I. Ion, R. Reeder, and S. Consolvo, ""...No one can hack my Mind": Comparing expert and Non-Expert security practices," in *Proc. SOUPS*. USENIX Association, 2015.

[89] W. Bai, M. Pearson, P. G. Kelley, and M. L. Mazurek, "Improving Non-Experts' Understanding of End-to-End Encryption: An Exploratory Study," in *EuroUSEC '20*, 2020.

[90] A. J. Reynolds, J. A. Temple, S.-R. Ou, D. L. Robertson, J. P. Mersky, J. W. Topitzes, and M. D. Niles, "Effects of a school-based, early childhood intervention on adult health and well-being: A 19-year follow-up of low-income families," *Archives of pediatrics & adolescent medicine*, 2007.

[91] F. Campbell, G. Conti, J. J. Heckman, S. H. Moon, R. Pinto, E. Pungello, and Y. Pan, "Early childhood investments substantially boost adult health," *Science*, 2014.

[92] E. B. Blinder, M. Chetty, J. Vitak, Z. Torok, S. Fessehazion, J. Yip, J. A. Fails, E. Bonsignore, and T. Clegg, "Evaluating the use of hypothetical 'would you rather' scenarios to discuss privacy and security concepts with children," *Proceedings of the ACM on Human-Computer Interaction*, no. CSCW1, 2024.

Appendices

A   Extended paper

The extended paper with is available at:
https://arxiv.org/pdf/2406.13017