Investigating Security Folklore: A Case Study on the *Tor over VPN* Phenomenon

MATTHIAS FASSL, CISPA Helmholtz Center for Information Security, Germany
ALEXANDER PONTICELLO, CISPA Helmholtz Center for Information Security, Germany
ADRIAN DABROWSKI, CISPA Helmholtz Center for Information Security, Germany
KATHARINA KROMBHOLZ, CISPA Helmholtz Center for Information Security, Germany

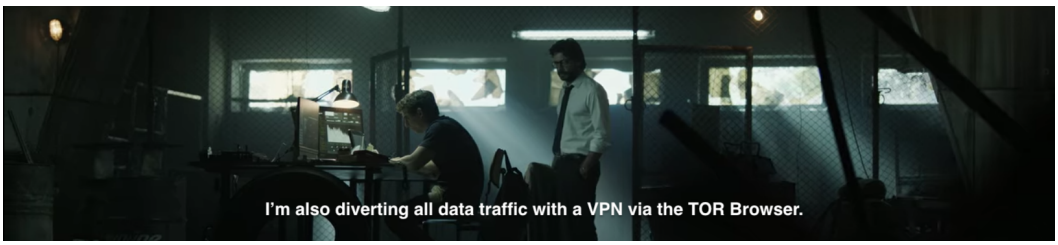


Fig. 1. Rio explains to the professor that he uses Tor over VPN to access the dark net (Money Heist S02E03, 2017 [50])

Users face security folklore in their daily lives in the form of security advice, myths, and word-of-mouth stories. Using a VPN to access the Tor network, i.e., Tor over VPN, is an interesting example of security folklore because of its inconclusive security benefits and its occurrence in pop-culture media.

Following the Theory of Reasoned Action, we investigated the phenomenon with three studies: (1) we quantified the behavior on real-world Tor traffic and measured a prevalence of 6.23%; (2) we surveyed users' intentions and beliefs, discovering that they try to protect themselves from the Tor network or increase their general security; and (3) we analyzed online information sources, suggesting that perceived norms and ease-of-use play a significant role while behavioral beliefs about the purpose and effect are less crucial in spreading security folklore. We discuss how to communicate security advice effectively and combat security misinformation and misconceptions.

CCS Concepts: • **Security and privacy → Human and societal aspects of security and privacy**; **Privacy protections**; **Pseudonymity, anonymity and untraceability**.

Additional Key Words and Phrases: Tor, VPN, Tor over VPN, Theory of Reasoned Action, Beliefs, Security Advice, Secure Experience

ACM Reference Format:
Matthias Fassl, Alexander Ponticello, Adrian Dabrowski, and Katharina Krombholz. 2023. Investigating Security Folklore: A Case Study on the *Tor over VPN* Phenomenon. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2, Article 344 (October 2023), 26 pages. https://doi.org/10.1145/3610193

Authors' addresses: Matthias Fassl, matthias.fassl@cispa.de, CISPA Helmholtz Center for Information Security, Germany; Alexander Ponticello, alexander.ponticello@cispa.de, CISPA Helmholtz Center for Information Security, Germany; Adrian Dabrowski, adrian.dabrowski@cispa.de, CISPA Helmholtz Center for Information Security, Germany; Katharina Krombholz, krombholz@cispa.de, CISPA Helmholtz Center for Information Security, Germany.

2573-0142/2023/10-ART344 $15.00
https://doi.org/10.1145/3610193

1 INTRODUCTION

An increasing number of laypeople are concerned about their privacy and security online and want to minimize their risks. They are met with a conundrum of possible products, advice, myths, and word-of-mouth stories. For example, some users append symbols to their passwords assuming increased security [67], believe that (unencrypted) SMS are more secure than email and encrypted instant messengers [2], assume that information on the blockchain is not readable by others despite it being a public ledger [38], or overestimate the privacy gains of private browsing features [1, 25, 30]. In these situations, users' folk theories [51, 68] about the purpose and effect of security behaviors do not reflect reality. People perpetuate *security folklore*, a collection of beliefs about security practices (discussed in more detail in Section 2.1) when they tell informal stories about their security experience to others [52]. At best, these behaviors do not impact privacy and security online. At worst, they give a false sense of security while negatively impacting security.

Currently, little is known about how security folklore becomes popular, spreads, and gets widely adopted. Understanding these mechanisms could facilitate the transport of accurate and up-to-date information about security practices. Of particular interest are factors that make them so easy to retell to others, intuitively understandable, feel so secure, and actionable in users' day-to-day life.

In this paper, we approach the broad topic of security folklore based on the peculiar case of *Tor over VPN*, where the user first connects to a VPN before connecting to Tor. Experts [65, 66] and special interest groups [56] doubt the general security benefits of this practice. The case of *Tor over VPN* is particularly interesting because its utility (or dangers) are actually not fully conclusive, it vividly illustrates the power of misconceptions, and it found entrance into popular culture (e.g., Figure 1). An in-depth understanding of the phenomenon, its mechanisms, and the forces at work may help us to start understanding the building blocks of security folklore.

We triangulate this phenomenon with three study parts that align with the *theory of reasoned action* [22]: (1) we measure the prevalence of the Tor over VPN practice on real-world Tor traffic, (2) we survey Tor over VPN users' beliefs about the practice using mixed qualitative/quantitative methods, and (3) we systematically analyze the online content of background information on Tor over VPN.

We found that Tor over VPN users, comprising 6.23% of daily Tor users, rarely articulated threat models and expected general security benefits. Users who gave reasons often wanted to protect themselves from the perceived dangers of using Tor. This attitude reflects how VPN providers tend to write about the practice, emphasizing and sometimes misrepresenting the risks of using Tor to recommend their VPN for accessing the dark net. Our document analysis suggests that background information establishes normative beliefs when users look to others for appropriate security behavior (i.e., social proof [11–13]), contributing to the spread of this kind of security folklore. The importance of normative beliefs, which do not rely on knowledge about the practice's purpose and effects, explains why many Tor over VPN users expect general security benefits. Establishing reliable behavioral beliefs about the effects of Tor over VPN is difficult since the information sources, mostly VPN providers and discussions in expert venues, often contradict each other. This work highlights that newspaper reports about security practices or demonstrated security behavior are specific forms of security advice – since they offer social proof for security behavior. Knowing the relevance of normative beliefs can improve the future dissemination of high-quality security advice. Reports about security practices must avoid unintentional misinformation and misconceptions. Always including the security practices' purpose and effects may help.

2 BACKGROUND AND RELATED WORK

We discuss the basics of security folklore, i.e., folk theories, security misconceptions, and security advice, before covering prior work about users' perceptions and use of VPNs and Tor. We include short descriptions of different Tor and VPN combinations and experts' opinions on the security benefits.

2.1 Security Folklore and Misconceptions

According to Brunvand's definition from 1978, folklore is a part of a culture that *"encompasses all knowledge, understandings, values, attitudes, assumptions, feelings, and beliefs transmitted in traditional forms by word of mouth or by customary examples"* [10]. When we transfer this notion to security and modern communication, security folklore is a collection of beliefs about security practices that have become part of the culture and are transmitted informally through, for example, verbal tales, written social media posts, or demonstrated behavior. This informal and repetitive communication may also enhance the perceived legitimacy of these security beliefs, regardless of available empirical evidence. In contrast to the repeated informal stories about similar security incidents that people tell each other, official, e.g., password-creation guidelines would not constitute security folklore.

Thus, there are two approaches to studying security folklore: investigating the informal communication about security practices and understanding users' collective knowledge, assumptions, feelings, and beliefs about security practices. In the spirit of the first approach, Rader et al. [52] and Pfeffer et al. [49] investigated the stories that people tell each other about security and what kind of lessons they learned from them. They found that people remembered these stories for months or years after initially hearing them and that almost half of them retold them to others. Studying users' functional or structural mental models [14, 36] of security technology would be a way to use the second approach to researching security folklore. For example, Wash [68] found folk models of home computer security that influence which security software people use and which expert security advice they follow. Rader and Slaker [51] found that these folk theories about technology depend on how the technology represents its actions to the users. Similarly, technology may provide a secure experience [41] not based on actual security. In this work, we investigate a specific case of security folklore in both ways, i.e., investigating users' beliefs and assumptions about Tor over VPN and studying different kinds of informal communication about it.

These folk theories about security mechanisms could contribute to users' security misconceptions, e.g., concerning private browsing [1, 25, 30], the security of electronic communication [2], secure password creation methods [67], or the anonymity of blockchain transactions [38]. Even when these security misconceptions have no direct negative influence on users' security, they can have side effects such as a security theater [61], i.e., when users feel more secure while the technology is not. This perception may lead to risk compensation behavior [60], i.e., users accepting more risks believing they are secure. In the face of limited compliance budgets [5], switching to better security practices is advisable.

2.2 Security Advice and Security Practice Adoption

Prior work [57, 59] suggests a large body of security advice from which users can choose. However, the quality of the security advice varies wildly. Redmiles et al. [58] analyzed a corpus of collected security advice along three dimensions: comprehensibility, perceived efficacy, and perceived actionability. They argue that users' limited compliance budget [5] necessitates a small set of high-quality security advice that benefits a broad range of people.

Understanding the circumstances that lead to user adoption of security and privacy measures is a matter of ongoing research. As Harborth et al. [31] found, perceived anonymity and trust in the service drive perceived usefulness and, consequently, adoption. Security practices requiring reoccurring user interaction are less widely adopted than those requiring fewer interactions [70]. The last factor might benefit VPN adoption since VPNs do not require a lot of user interaction compared to other security mechanisms. Also, the theory of reasoned action [22] can help to understand security practice adoption. It provides an overview of how background factors influence different beliefs about practices, how these beliefs affect user intention, and when this intention translates to user behavior.

2.3 The VPN Ecosystem and its Users

Commercial VPN providers are a 15-billion dollar industry that markets their services as a turnkey solution to users' privacy [37]. The industry invests heavily in marketing. Akgul et al. [4] investigated influencer marketing ads on YouTube for VPNs. They found misleading claims, including overpromises and exaggerations, that could negatively influence users' mental model of internet safety. Ads like these may increase users' trust in the services' offered privacy and lead to their adoption [31]. Ramesh et al. [54] found that security and privacy are people's main reasons for adopting VPNs, while around 40% of their participants have flawed mental models of them. They attribute aggressive and misleading VPN ad campaigns to the degradation of users' mental models.

Binkhorst et al. [6] found that non-experts and experts alike were unsure when to use a VPN. They partially explained this with both groups' unclear threat models. Users adopt and abandon VPNs for emotional and practical reasons [45], whereas university students adopt them primarily for practical reasons [18].

2.4 Combined Tor and VPN Use: Potential Benefits and Harms

In general, non-expert Tor users employ abstractions that hide essential operational aspects of Tor [24]. These abstractions may lead to behavior that compromises non-experts' anonymity.

For example, Biryukov and Pustogarov [7] discussed the anonymity effects of combining Tor with Bitcoin. Some developers endorsed Tor for Bitcoin to avoid IP address leaks. However, this only provides limited anonymity and introduces an additional attack surface for MitM attacks. Story et al. [62] studied the adoption of security tools and found that users with VPN and Tor browser experience were often confused about their protection. Several of their participants stated that the Tor browser must be used with a VPN for added security. There are several approaches to combining Tor with a VPN. The two common ways are Tor over VPN, i.e., connecting to a VPN before opening the Tor browser, and VPN providers' Tor mode, where VPN servers relay the users' traffic into the Tor network so that users do not need to install the Tor browser. Other niche approaches, such as using a VPN over Tor or elaborate combinations of different tunneling technologies, exist but are not the focus of this work.

While experts doubt the general security benefits of Tor over VPN, they agree on two specific cases where it is helpful: Circumventing censorship when access to the Tor network is blocked while access to VPNs is not, and, to some extent, hiding Tor traffic from ISPs [56, 65, 66]. However, Tor bridges may provide similar benefits in both cases [66].

Since VPN providers are Tor over VPN users' first hop forever, the same concerns as with potentially malicious Tor guard nodes apply: they are in the position for selective denial of service attacks or statistical profiling attacks [16]. Either could aid attackers in deanonymizing Tor users linking them to their online behavior. Hence, Tor over VPN users must trust their choice of VPN provider completely [15, 66]. However, as prior work [33, 53, 69] has noted, statements by VPN providers are not necessarily trustworthy and using VPN may do more harm than good.

Measuring traffic in the Tor network in a privacy-preserving manner has been an active research topic in the last few years, resulting in PrivCount [34] and PSC [20, 39]. Our measurement study builds on PrivCount to quantify VPN use in the Tor network.

3 METHODOLOGY

In this case study, we investigate the phenomenon of Tor over VPN practice to learn more about the mechanisms of security folklore. Hence, we are not only interested in the efficacy of Tor over VPN (object) and the users who practice it (subject) but also in the *"terms, conditions, and situation of the interaction"* [29] between them. Initially, we want to know if Tor over VPN practices are widespread enough to consider them part of the culture. Then, we investigate this part of security folklore in two ways: first, by studying users' assumptions and beliefs about the practice, and second by investigating informal communication about the practice. Hence, we focus on the following three research questions:

RQ1: How widespread is the behavior of using Tor over VPN amongst Tor users?
RQ2: What are the users' reasons for employing Tor over VPN?
RQ3: Which communication mechanisms contribute to diffusing Tor over VPN information and behavior?

To guide our investigation of this security practice, we align our case study with the theory of reasoned action [22]. The theory of reasoned action postulates that human behavior, while not necessarily rational, follows basic patterns. The authors proposed this model to understand the determinants of behavior and, in the long run, design behavioral interventions. According to the theory, behavioral intention and actual control are strong indicators that individuals will carry out a behavior. These intentions are based on three belief types: behavioral, normative, and control.

Behavioral beliefs are about the expected positive or negative consequences of performing a certain behavior [22, p. 100]. The intention to perform the behavior is primarily influenced by possible outcomes that individuals perceive as probable and vital.

Normative beliefs, in general, are the foundation for "perceived social pressure to perform a given behavior" [22, p. 130]. Fishbein and Icek [22, p. 131] differentiate two types of norms: injunctive norms, which concern other individuals' or groups' moral judgment on behavior, and descriptive norms, which concern the perception of other individuals' or groups' behavior. Related Usable Security research [12, 13] also uses the term *social proof* for these descriptive norms to describe how people look to others to learn the appropriate behavior.

Control beliefs form the "perception that one has or does not have the ability to carry out the behavior (i.e., perceived behavioral control)" [22, p. 170]. These beliefs are based on several control factors, such as required resources, available opportunities, anticipated obstacles, and self-efficacy.

Combined, these three beliefs influence peoples' intention to perform the behavior in question. The last belief about self-efficacy moderates the intention to engage, i.e., people only plan to implement a behavior when they are confident that they have the necessary abilities. These beliefs may originate from diverse sources, e.g., formal education, prior experiences, newspapers, TV, other Internet media, or interactions with friends and family. Other background factors like personality and demographics can influence how people interpret and remember information about the considered behavior. In contrast to other behavioral theories, the theory of reasoned action also incorporates background factors that inform these behavioral, normative, and control beliefs. Since this theory provides a unified framework for social behavior, it helps us guide our investigation and find the determinants of the Tor over VPN practice.

Hence, we investigate the Tor over VPN phenomenon in three separate studies, each focusing on a different aspect of the theory of reasoned action (as shown in Figure 2):

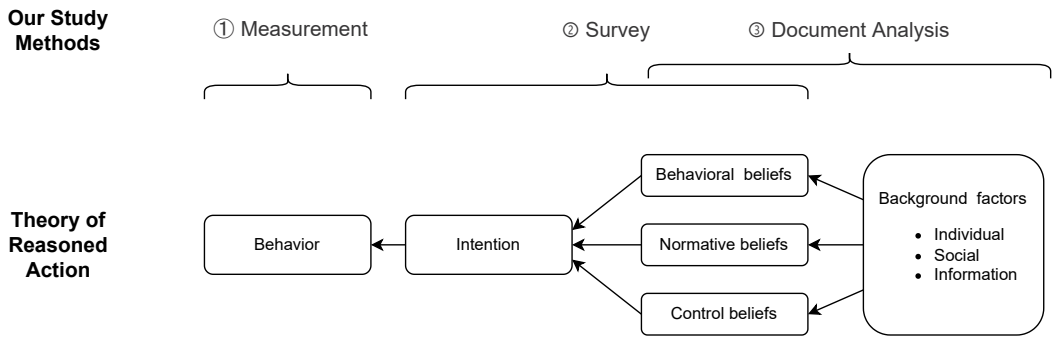


Fig. 2. Our study methods' alignment with the theory of reasoned action

(1) we use a privacy-preserving method to *measure Tor over VPN behavior* in the Tor network to understand how widespread this practice is (see Section 4 and RQ1),
(2) we use Amazon MTurk to *survey Tor over VPN users' beliefs* about the practice to understand more about why they choose to implement this practice (see Section 5and RQ2), and
(3) we apply *document analysis to online information sources* covering Tor over VPN to understand on which information users may base their beliefs (see Section 6 and RQ3).

Since we also consider discussions on social media in our document analysis, we inadvertently also analyze user beliefs. Figure 2 illustrates this overlap between these two study parts with brackets. The upcoming sections describe these three studies in detail, including the methods, ethical considerations, and results.

4 MEASUREMENT OF TOR OVER VPN BEHAVIOR

To study the prevalence of VPN use in the Tor network, we set up four guard nodes to measure the incoming traffic from detected VPN endpoints. Using this approach, we measure how human behavior impacts the entire Tor network in real-world conditions. However, we cannot directly conclude the number of VPN users in the Tor network from our measurement results since users might share the same VPN endpoints.

In the remainder of this section, we describe the measurement setup, the VPN decision procedure, counting Tor traffic data in a privacy-preserving manner, and analyzing the results.

4.1 Measurement Setup

We set up four Tor relays in a German data center. We started our measurement after the relays received their guard flag (showing that the relays had been around at least eight days[1]) and reached their bandwidth capacity. At this point, they advertised a combined total bandwidth of 1.46Gb/s. The probability of choosing one of our guard nodes was 0.6%, which made them the ninth-largest group of guards in the Tor network at the time, according to OrNetStats [47].

We use PrivCount [34], a system for safely measuring the Tor network, to count incoming connections' parameters. As Figure 3 shows, our PrivCount setup includes four data collectors (one for each guard node), two share keepers, and one tally server. Guard nodes ran a PrivCount-patched version of Tor, which notified the corresponding data collector of all connection events. Since the PrivCount patch was only available for an outdated version of Tor, we ported it to a

[1] https://blog.torproject.org/lifecycle-new-relay

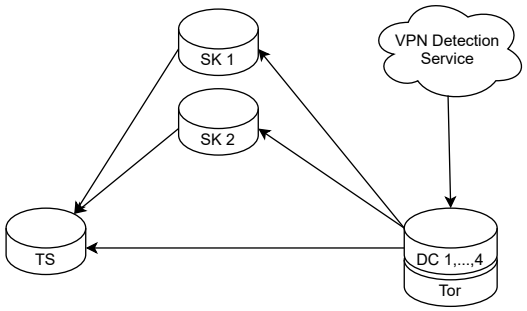


Fig. 3. Measurement setup containing four data collectors (DC 1-4), two share keepers (SK 1-2), and one tally server (TS). Each data collector receives events from a Tor guard node and retrieves information from a VPN detection service.

supported version. Then, we adapted the data collectors to count the incoming connections from VPN providers.

4.2 VPN Decision Procedure

Commercial VPN services advertise their product as a way to access blocked content [37]. In response, content providers are interested in detecting VPN or proxy connections, which creates a market for so-called fraud protection services. Our initial market research uncovered the following services: proxycheck.io, IPtrooper, IPQualityScore, getipintel.net, IPHub.info, IPHunter, vpnblocker, IP2Location, Shodan, IPWarner, and MaxMind. All of them offer an online API service. Three of them, IPQualityScore, IP2Location, and Maxmind, also provide a privacy-preserving alternative: downloading a regularly updated database. We pre-filtered these services using basic quality criteria:

(1) *Specific responses*: a service's response must indicate if the IP address belongs to a VPN.
(2) *Rate limit:* services need to either support a minimum of 20,000 queries per day (i.e., the expected daily amount of connections per guard node) or provide an option to download a database for offline queries.

To select our VPN decision procedure, we evaluated the accuracy of the remaining services that fulfill these two criteria. We collected a ground-truth dataset consisting of known VPN and non-VPN addresses. For the set of known VPN IP addresses, we created accounts at four popular VPN providers (ExpressVPN, HideMyAss, NordVPN, and PIA), connected to all VPN endpoints accessible to us, and retrieved each endpoint's IP address (487 in total). For the set of known non-VPN IP addresses, we fetched a similar amount of sites from Alexa Top Sites and resolved them to IP addresses (485 in total). We consider these commercial sites suitable for testing a VPN decision procedure since VPN endpoints, similar to commercial sites, are commonly located in data centers rather than residential IP ranges. We included the complete dataset in the supplemental material.

To compare the quality of these services, we calculated Recall, Precision, and the F1 score. The F1 score combines Recall and Precision equally and measures classifiers' accuracy. Calculating the precision and the F1 score requires careful consideration since they depend on the prevalence of the underlying category (VPNs), which is what we want to measure in the first place. Therefore, we assumed a prevalence of 0.05 for this part of the evaluation and adjusted our balanced dataset accordingly.

4.3 Differential Privacy

Like Jansen and Johnson [34], we decided on a 24-hour measurement period in all cases. Increased measurement periods do not reliably improve the accuracy since the differential privacy approach adds the corresponding amount of noise, and the set of guard users only changes slowly. Our data collectors receive information about each connection's byte count, circuit count, origin IP address, and the number of concurrent connections. Using this available data, we count the incoming connections, the amount of transferred data, and the number of circuits per connection, once for VPNs and once for all incoming traffic.

Our connection-based measurement approach does not allow differentiating between different users' connection data. Instead, we establish action bounds that define the privacy that our collection affords. We protect 24 hours of continuous use, including one new entry connection every other hour, six new circuits per hour (the typical circuit lifetime is 10 minutes), and two preemptive circuits. We also protect 10MiB of traffic in either direction. Since we did not know how many counts per hour we could expect, we conducted short pilot measurements and extrapolated these counts to 24 hours. Similar to Jansen and Johnson [34], we use the differential-privacy parameters of $\epsilon = 0.3$ and $\delta = 10^{-6}$ for our PrivCount deployment. We used the defined action bounds, the expected counts, and the differential-privacy parameters to calculate the appropriate noise levels with the noise computation script included in the PrivCount project.

4.4 Analysis

We analyze the measurement results by adjusting all the VPN traffic counts, calculating the 95% confidence intervals, and calculating the share that VPN traffic contributes to the overall counts.

Foreman [23] describes several methods that accurately count positives with inaccurate classifiers. One of the methods he describes is the "Adjusted Count" shown in Equation 1. This method only requires knowing the true-positive (TPR) and false-positive rates (FPR). These are independent of the prevalence, so we derived them using our measurements from Table 1.

$$adjusted = \frac{observed - FPR * total}{TPR - FPR} \tag{1}$$

Using this method, we adjust all measurements of VPN traffic before analyzing them further. We report both the observed and the adjusted measurements.

PrivCount adds noise to all counter values. This noise has a normal distribution, with a mean of zero and a pre-calculated standard deviation (see Section 2.3 in Jansen and Johnson [34]). We compute the 95% confidence interval to report the uncertainty that the added noise introduces.

When we calculated the VPN traffic's share of the overall counts, we used interval calculus operations to calculate the resulting intervals. These operations increase the range of the 95% confidence intervals for these measurements.

To ensure replicability and facilitate research cooperation, we will publish the scripts we used to evaluate the VPN detection services, the PrivCount-patched Tor version, our modified PrivCount project, and the evaluation scripts on Github.

We present the results of our measurement study in three parts: (a) the evaluation results of the VPN detection services in Section 4.6, (b) a comparison of VPN and non-VPN traffic patterns to check for systemic biases in Section 4.7, and (c) the measurement of the prevalence of VPN use in the Tor network in Section 4.8.

4.5 Ethical Considerations and Limitations

Before collecting data, we submitted a description of our motivation and methodology to the Tor Research Safety Board (TRSB). We adapted our method according to the feedback we received. Furthermore, our institution's ethical research board (ERB) approved this study.

In this measurement study, we face two ethical challenges: (1) Participants could not consent to the data collection since Tor clients choose guard nodes randomly (weighted by bandwidth). To mitigate this lack of information, the domain name of the Tor relays indicated their use for research purposes. In the Tor browser, clicking the lock icon in the address bar shows the currently used Tor circuit and addresses of the individual Tor nodes. Accessing the Tor node address with a web browser led to a website that explained the purpose of our research and how users could opt out of participation. However, since we did not log access to this website and had no feasible option to measure how many users opted out, we can not say how this affected the validity of our measurement. (2) The measurement procedure handled sensitive data, namely the IP address and access time of connecting users. The measurement setup did not store these data points but discarded them after increasing the corresponding counters. After careful evaluation, we checked for VPNs with a local version of IP2Location's database. Hence, the users' IP addresses did not leave the guard node. To ensure our criteria for handling sensitive data, we only used Tor nodes under our control and did not collaborate with other established Tor guard families. We did not retain or share users' IP addresses and time of connection at any point. Neither the Tor daemon nor any other service logged the incoming connections on these Tor relays. Similar to prior work [34], we employed differential privacy to protect the data of individual users. Attackers cannot tie data that leaves the Tor measurement nodes to specific users as long as their traffic stays in our protected action bounds. The main risk associated with our data collection is an inference attack on the behavior of specific users of our Tor nodes. Our differential privacy approach mitigates this risk in case the users remain inside our defined action bounds.

All of our measurement nodes were located in Germany. If a relevant number of Tor users opted out of using Tor nodes in Germany, this may have impacted our sample's quality. However, Germany is the most common location for Tor nodes, with around a fifth of all Tor nodes, and the data center we used is the second most popular data center for Tor nodes. Hence, opting out of Tor nodes from the data center we used or Germany as a country would result in significantly fewer available Tor nodes, impacting the overall user experience. User behavior changes over time, and current events affect them. Hence, repeating the measurement of VPN use in the Tor network may yield different results. While we can measure behavior, the measurement data does not tell us about users' reasons for their behavior. Hence, we conduct a user survey and a document analysis to complement the measurement results (see Section 5).

4.6 Evaluation of VPN Detection Services

Our market research of existing VPN detection services resulted in five different candidates. We evaluated them with known VPN and non-VPN addresses as described in Section 4.2. We calculated recall, precision, and the F1 score for each service. We assumed a prevalence of 0.05 for the calculation of the latter two. Table 1 shows the results of this evaluation.

The low precision of most services in the list at a low prevalence would overestimate the amount of VPN traffic in our measurement. Only Shodan and IP2Location provide suitable precision. However, since Shodan relies on manually labeled datasets, the recall is very low. Even though we prioritize these services' precision, we also look at the F1 score, which combines recall and precision. IP2Location's service achieves our dataset's best F1 score (0.87). Therefore, we used the IP2Location

Table 1. Evaluation of selected VPN detection services

	PC	IPQS	Shodan	MM	IP2L
True Positive	456	474	106	482	404
False Negative	31	13	381	5	83
True Negative	353	94	485	306	483
False Positive	132	391	0	179	2
Recall	0.94	0.97	0.22	0.99	0.83
Precision*	0.15	0.06	1.00	0.12	0.91
F1*	0.26	0.11	0.36	0.22	0.87

* Assumes a VPN prevalence of 0.05
VPN detection services: Proxycheck (PC), IPQualityScore (IPQS), Shodan, Maxmind (MM), and IP2Location (IP2L)

database, which our measurement setup downloaded every time before starting a measurement. We improved its count accuracy with the adjustment method described by Foreman [23].

4.7 Validating (Non-)VPN Traffic Classification

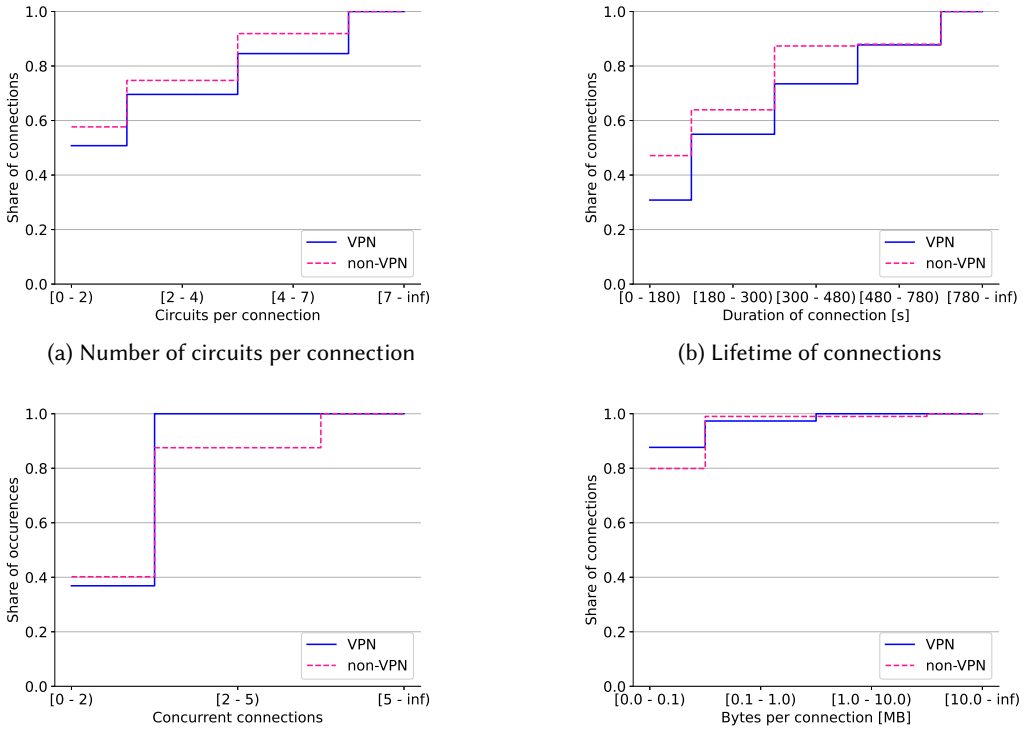
We compared the VPN and non-VPN traffic in our measurement period to further validate our VPN classification procedure. We expect that VPN users' traffic follows the same patterns as non-VPN users' traffic. Significant differences in traffic patterns suggest a systemic bias in our classification procedure.

We collected the following information for each connection: lifetime, amount of circuits, concurrent connections, and the number of transferred bytes. The cumulative histograms in Figure 4 highlight the similarities and differences in these data points between the VPN users (solid blue lines) and non-VPN users (dashed red lines). Since the base rate of both groups differed in our dataset, we scaled the results along the y-axis for easier comparison. Subfigures (a), (c), and (d) in Figure 4 show similar VPN and non-VPN traffic in our measurement period. Increasing our confidence that the classification does not introduce major systematic bias.

Subfigure (b), a cumulative histogram of the connections' lifetime, is the only one that shows some differences. While the number of VPN connections monotonously decreases with increasing lifetime, the non-VPN connections' lifetime shows two spikes. The first spike is between 0 and 3 minutes, which is interesting since Tor should keep connections open for at least 3 minutes, even in cases without activity [17]. The second spike appears at five minutes. We have no conclusive explanation for these differences in connection lifetime between VPN and non-VPN traffic. Another interesting detail is that, for both VPN and non-VPN traffic, many connections only have one circuit. However, according to the Tor path specification [17], Tor clients aim to maintain two pre-built circuits for each recently seen port.

4.8 Measured Tor over VPN Prevalence

Using our setup, we measured three data types: the number of connections, circuits, and transferred data. We separate these data types into two bins, the total amount and the amount coming from detected VPN endpoints. Table 2 shows an overview of these measurement results. Next to each reported measurement, we added the 95% confidence interval in brackets. We report the adjusted counts [23] as well as the observed counts. Additionally, we calculate the ratio of VPN counts to



(a) Number of circuits per connection

(b) Lifetime of connections

(c) Number of concurrent connections

(d) Transferred data per connection

Fig. 4. Cumulative step histograms that compare VPN (solid blue line) and non-VPN (dashed red line) connections

Table 2. PrivCount measurement results for connections, transferred data, and circuits before and after adjusting the counts.

	Connections (in thousands)		Transferred Data (GiB)		Circuits (in thousands)	
Total	10 186.9	[9 662.4, 10 711.4]	3 039.3	[2 984.7, 3 094.0]	42 952.5	[42 092.1, 43 812.8]
Count VPN						
observed	565.6	[546.8, 5 843.9]	177.7	[174.6, 180.9]	2 533.4	[2 501.7, 2 565.2]
adjusted	634.3	[615.5, 653.1]	200.2	[197.0, 203.3]	2 854.6	[2 822.8, 2 886.3]
Percentage VPN						
observed	5.55	[5.10, 6.05]	5.85	[5.64, 6.06]	5.90	[5.71, 6.09]
adjusted	6.23	[5.75, 6.76]	6.59	[6.37, 6.81]	6.65	[6.44, 6.86]

Note: The brackets next to each result include the 95% confidence interval.

total counts for all measurement variables. We used interval calculus operations to determine the resulting intervals for the calculated shares.

As Table 2 shows, our family of four guard nodes handled over 10 million connections, 3 TiB of data, and 43 million circuits in the 24-hour-long measurement period. A total of 6.23% of the connections, 6.59% of the transferred data, and 6.65% of the circuits originate from VPNs. The

95% confidence interval is < 1% in all three cases. These traffic measurements include all different ways of using a VPN to access the Tor network, i.e., regular Tor Browser users and users of VPN providers' Tor mode as described in Section 2. These results confirm that the practice of accessing the Tor network via a VPN provider is comparatively widespread, considering its doubtful benefits [56, 65, 66].

5 SURVEY OF USERS' TOR OVER VPN INTENTIONS AND BELIEFS

In the context of the theory of reasoned action (see Figure 2), our measurement approach shed light on Tor users' behavior but did not help us understand their intentions, beliefs, or the background factors that led to them. For example, users may always use a VPN or explicitly apply Tor over VPN as a security practice regardless of context. Hence, we continue our investigation with a survey to understand Tor users' intentions and initial beliefs behind the behavior.

To investigate, we initially searched for prior work that provides insights into VPN usage patterns. According to the Global VPN Usage Report 2020 [64], 31% of internet users have used a VPN in the month before the survey. Of these VPN users, 36% reported using a VPN on a (nearly) daily basis. The report does not provide details on general VPN use that does not consider situation-specific context.However, the report states that accessing the Tor Browser is a reason for VPN use for 14% of VPN users. While this does not explain how much of our measurement results can be explained by these users, it confirms Story et al.'s [62] observation that some users intentionally apply 'Tor over VPN.' We conducted an online survey to help us interpret our measurement results and understand users' reasons for combining Tor with VPNs.

5.1 Method

We recruited $N = 119$ participants on Amazon MTurk. To learn more about Tor over VPN behavior and perceptions, we focused our recruitment on participants who had used both the Tor Browser and VPNs but not necessarily in combination. This results in a narrower sample than the previous measurement study, where we sampled real-world Tor users corresponding to our measurement nodes' bandwidth ratio.

We wanted to recruit participants with VPN and the Tor Browser experience. Only around one-third of Internet users seem to use VPNs at least occasionally [64] and definitive numbers for the prevalence of Tor Browser use are hard to find. Consequently, we were concerned about how many participants with that experience we would be able to recruit. As Kang et al. [35] showed, MTurk workers are more privacy-conscious than the U.S. public. Hence, we chose it for this survey because we hoped to reach our targeted population better on MTurk. However, as Tang et al. [63] showed after we conducted the survey, the same is true for Prolific. Since they also found that samples from Prolific or CloudResearch offer increased data quality, we would use one of them for future work on the Tor over VPN phenomenon.

The supplemental material contains the complete questionnaire. We asked if the participants used a VPN, whether they accessed VPNs for specific purposes or regardless of context, and whether they preferred using a VPN to access the Tor network or not. For the latter two questions, we included open-ended questions asking participants to explain their preferences.

We evaluated the open-ended questions using thematic analysis [9]. One author performed an open-coding step on all responses to construct one codebook per question. The supplemental material contains all resulting codebooks. Another author used these codebooks to code the responses independently. During the coding process, both authors noted potential themes in the data. The initial inter-coder agreement Krippendorff's alpha was $\alpha = .59$ and $\alpha = .64$ for the first and second questions, respectively. Both authors met and discussed conflicts while updating the codebooks as necessary. The discussion increased the inter-coder agreement Krippendorff's alpha

Table 3. Cross-tabulation of the participants' preferred way of accessing the Tor network and their VPN usage pattern.

		Tor Access Preference			
		VPN	no pref.	no VPN	$\sum$
VPN usage	Specific	42	3	9	54
	Usually	18	2	1	21
	$\sum$	60	5	10	75

to $\alpha = .89$ and $\alpha = .94$. Finally, both authors discussed their notes and identified themes in the data. We present these themes in Section 5.4.

We required that participating workers maintain an approval rate of at least 95% for their past MTurk tasks. We received data from 190 MTurk workers. They spent an average of 3 minutes and 52 seconds on the survey, and we compensated all of them with USD 0.62. We excluded data from workers that have not used Tor (31) or failed the attention check question (40), resulting in a final dataset with responses from 119 participants. The median age of the included participants was 30 ($std = 7.95$). Overall, 27 (23%) women and 92 (77%) men participated. No participant selected one of the remaining response options, i.e., 'non-binary,' 'I prefer not to say,' or 'I prefer to self-describe.'

5.2 Ethical Considerations and Limitations

For our belief survey, we informed all participants about the study purpose and data handling practices on the first page of our survey. We compensated all survey participants and did not ask questions about their use case for the Tor browser. Per Amazon MTurk's guidelines, we did not collect any personally identifiable data. Our institution's ethical research board (ERB) approved this study.

Tor users prefer to stay anonymous. Recruiting Tor over VPN users for a survey that questions them about their observed behavior on the Tor network is ethically and technically challenging. For our survey on Tor over VPN beliefs, we recruit users on Amazon MTurk who self-report experience with Tor as well as VPNs.

5.3 (Un)intentional Use of Tor over VPN

Seventy-five (63%) of the 119 recruited Tor users had a VPN client installed on one of their devices. Table 3 shows these 75 participants preferred way of accessing the Tor network and VPN usage pattern. Sixty participants preferred to access the Tor network using a VPN. Of these, 42 (70%) reported using a VPN for specific situations, and 18 (30%) reported using a VPN regardless of context.

Fundamentally, there are two types of Tor over VPN users: (1) those who use a VPN regardless of context and happen to use Tor at the moment, and (2) those who specifically connect to a VPN before accessing the Tor network. Our survey results suggest that the latter type of intentional user is responsible for most of the VPN traffic in the Tor network.

5.4 Tor over VPN as a Security Practice

Most participants who used a VPN regardless of context, reported doing so for the general benefit of added security. They expected an extra layer of protection to guard their privacy online, with some stating that their VPN software provided them with a more secure experience. Participants connecting to VPN only occasionally named more diverse motivations for doing so: The most

prominent reason was to bypass geo-blocking or otherwise hide their location from a service. Participants also used VPNs to hide other personally identifiable information (PII), such as IP addresses, from services. Others used VPN in untrusted networks, such as public WiFi, or due to work requirements. Finally, several participants explained that they connect to a VPN to hide specific activities from observers, most notably their internet provider. These include legal gray areas such as filesharing or torrenting. Three participants, who reported using VPNs for legal gray areas, explicitly named accessing the Tor network as a reason to connect to a VPN. Users seem to associate Tor with activities they would rather hide from internet providers and other observers.

We attributed users' reasons for combining Tor and VPN to behavioral, normative, or control beliefs. When participants mentioned concrete threat models or protection mechanisms, we treated them as behavioral beliefs. When trust issues and others' recommendations were prime concerns, we treated them as normative beliefs. When participants focused on the action itself, i.e., that it is one more thing they could do or how it made them feel, we classified this as a control belief.

Behavioral Beliefs. When investigating users' Tor access preferences (with or without a VPN), we found eleven participants who used a VPN to protect themselves from threats they identified within the Tor network. Some knew that guard nodes could collect their IP address and used a VPN to conceal this information. Participants associated the Tor network with unlawful activities, expected to find *"unsavory characters"* (P102), or suspected a trap by governmental agency: *"I have read some places that Tor was set up by the FBI."* (P95). Therefore, they employed extra protective measures to preserve their privacy. Other participants stated that, due to the open nature of the Tor network, i.e., everyone can contribute to the open-source project and run Tor nodes, malicious actors could easily infiltrate the Tor network and collect information about its users. Similarly, P37 pointed to past incidents affecting the Tor network to explain their distrust: *"There was a problem they did not know about, and the user's information may have been shared by someone else."* Finally, some participants named uncertainty about the legality of Tor usage (in their country) as a reason to use a VPN before connecting to the guard node, hoping to evade prosecution. However, no participants brought up Tor bridges (in this context and throughout our study).

Normative Beliefs. Aside from the motivations mentioned above, we found that most participants, who preferred using a VPN when accessing Tor, referred to general security benefits as their main reason. These findings suggest that participants trust their VPN provider more than the Tor network. This difference in trust, even if rooted in wrong assumptions, can be hard to overcome. One possible explanation may be that users actively choose their VPN provider, while Tor randomly assigns guard nodes by default. Therefore, the observed trust might be related to choice-supportive bias, as described by Mather and Johnson [40]. Humans tend to attribute significantly more positive features to options they consciously chose and justify past decisions to themselves by perceiving their choice as superior to other options. Finally, a few participants explained that they adopted the practice of combining a VPN with Tor after getting recommendations from friends or other trusted sources. P74 disclosed: *"After conducting some research before starting to use it, I found it improves your privacy and security by using a VPN to access Tor."*

Control Beliefs. Participants often displayed a linear additive perception of security, where more measures equal improved security. P73 stated: *"If I'm trying to be private about what I'm doing I might as well set up as many safeguards as I can."* This assumption might hold poorly when combining Tor with a VPN, as we examined in Section 2. Other participants explained that combining these two technologies made them feel more secure. If such a perception does not coincide with actual security guarantees, users might take increased risks while navigating the web [60].

To summarize, we found that some (11) participants had concrete threat models for using Tor over VPN, such as surveillance by their internet provider or guard nodes, legal uncertainty, or mistrust in the Tor network. However, other participants employed Tor over VPN because they believed in linear additive security (i.e., the more, the better), followed others' recommendations, or simply because they felt more secure that way.

6 ANALYZING TOR OVER VPN INFORMATION SOURCES

In the previous Section, we surveyed users' beliefs about the Tor over VPN practice. According to the theory of reasoned action, several background factors influence these beliefs. An important factor, especially for security and privacy-motivated users, is the available information on the topic. Thus, analyzing available information about Tor over VPN may allow us to understand how it affects users' beliefs.

According to Redmiles et al. [57], users often learn about security behavior from news articles (online, print, and TV), online forums, fictional narratives, digital service providers, and advertisements. Hence, we focused on these sources for our document analysis of available Tor over VPN information.

6.1 Method

We used a keyword search to acquire an initial corpus of documents, filtered search results to end up with topic-relevant corpus entries, and then analyzed their content.

Creating an initial corpus. After an initial discussion, we acquired documents from online newspapers, VPN service providers, and social media, i.e., Twitter, Reddit, and Stack Exchange. For newspapers, we focused on the 16 most-read online newspapers in the U.S. and Germany [46] since all involved researchers understand them well. We investigated the 20 most popular providers [44]. Since the market for VPN services is highly concentrated [37], these providers likely cover a significant portion of users.

We employed the keywords "Tor VPN" and "Onion VPN" to find relevant content. We used a site-specific Google search query (e.g., "site:mullvad.net Tor VPN") to find relevant newspaper articles and service provider sites. We used the built-in search feature on Reddit and Stack Exchange. For the online newspapers and VPN providers, we included all results on the first Google result page in the sample. For Reddit and Stack Exchange, we sampled the first 200 results for each search query. For Twitter, we used Atlas.ti's built-in Twitter import feature, which limits data collection to the previous seven days. This procedure resulted in an initial corpus of 1015 entries, consisting of 249 newspaper articles, 82 VPN provider sites, 257 threads on Reddit, 310 threads on Stack Exchange, and 117 tweets.

Filtering relevant information sources. The entries in the initial corpus contained material where the terms Tor and VPN just happen to coincide in the same document. However, we were only interested in entries that cover the practice of combining Tor with VPN. Hence, we inspected and filtered all search results. Two researchers independently classified the first 110 corpus entries (i.e., 10.8%), comprising 82 VPN providers pages and 28 threads on Stack Exchange. They made the identical exclusion decision in 92.7% of the cases, corresponding to a Krippendorff's alpha of $\alpha = .87$. Since both researchers agreed on how to apply the exclusion criteria, one of them applied the criteria to the rest of the corpus. This filtering operation resulted in a corpus of 389 entries, consisting of 42 newspaper articles, 47 VPN provider sites, 110 threads on Reddit, 133 threads on Stack Exchange, and 57 tweets. Some entries were ambiguous and allowed for different interpretations, e.g., it was unclear if the article authors recommended using Tor and VPNs separately for different use cases or the two in combination. Both researchers discussed the issue and decided that the readers'

potential perception is more relevant than the authors' intentions for our analysis. Thus, the corpus should include these ambiguous cases. During the filtering process, the researchers kept notes on interesting observations and analytic thoughts, which they used as a starting point for the qualitative analysis.

Qualitative analysis. We combined deductive and inductive approaches to code the corpus entries. Because of our previous study part and our alignment with the theory of reasoned action, we deductively focused on information that may appeal to consumers' behavioral, normative, and control beliefs. Hence, explanations about the supposed effects of combining Tor over VPN that include threat models would align with behavioral beliefs, statements about who is or should use Tor over VPN contribute to normative beliefs, and justifications about the ease of use align with control beliefs. However, we also applied inductive open coding to other types of information, e.g., conceptions of technology, when we found them relevant to the analysis of the Tor over VPN practice. As Ortloff et al. [48] recommend for qualitative research in security, we describe the involved researchers' level of expertise: Two early-career researchers conducted this qualitative analysis of information sources. Both hold a Master's degree in computer science, have a strong background in security, and prior published research focusing on qualitative analysis. The two researchers started by cooperatively coding 48 documents, twelve each from VPN provider pages, newspaper articles, Reddit posts, and Stack Exchange questions, introducing codes as necessary. Afterward, they discussed the initial codebook and used categories to structure it. Using this initial codebook, both researchers independently coded the rest of the documents in three sessions, comprising 48, 50, and 188 documents, respectively. Both researchers kept notes about general observations and annotated quotes when they implied new codes or concepts. After each session, they discussed and resolved disagreements, adapting the codebook as necessary. In order to identify relevant areas and topics for discussion, the researchers calculated inter-coder agreement after each step, both across document groups and code categories [42]. The resulting inter-coder agreement values aided in adapting the coding process, if necessary. Most disagreements arose because the semantic meaning of the assigned codes was quite close or because the underlying data was ambiguous and allowed different interpretations by readers, including consumers seeking information. In the first individual coding session, we assigned codes to entire documents, which resulted in low inter-coder agreement values for detailed discussion threads on Stack Exchange and Reddit. Therefore, one researcher segmented these documents before the second and third individual coding sessions. In the second individual coding session, we found almost no new concepts, suggesting we had reached data saturation with our chosen information sources. Thus, we coded the rest of the documents in the third individual coding session. Since both researchers coded the entire corpus and agreed on the established concepts through discussion, it was optional to calculate an overall inter-coder agreement value [42]. Finally, the researchers discussed how the analyzed data answers the research questions and how to structure and present the findings. The supplemental material contains the final codebook and the complete filtered corpus.

6.2 Ethical Considerations and Limitations

For our document analysis, we collected and analyzed users' comments on Reddit, Stack Exchange, and Twitter. We considered it infeasible to collect informed consent from all of them and filter the research data accordingly. In line with Gilbert et al. [27] recommendations, we try to adhere to the context-dependent norms and platform affordances of Twitter, Reddit, and StackExchange: We did not collect or use any limited-visibility communication and did not try to predict any identities from collected site data. On the contrary, we protect site users (except for well-known public figures) by not analyzing, reporting, or quoting any information that could potentially identify them – which

is in line with the recommendations by Fiesler and Proferes [21] on research with Twitter data. Our institution's ethical research board (ERB) approved this study.

Our document analysis considers document types ranging from newspapers to social media channels. However, information consumption behaviors vary among demographic groups, leading to different discussions and presented information. Hence, including further information sources may lead to different analysis outcomes.

6.3 How Tor over VPN Information Sources Affect Beliefs

Overview of the filtered corpus. We begin with an overview of the characteristics of data in our filtered corpus to provide context for our document analysis. Our corpus spans the period from 2010 to 2022. Renowned newspapers, e.g., The New York Times, The Washington Post, or NPR, write about Tor and VPNs in the context of current events. For example, ongoing censorship in Russia during the war in Ukraine, access to abortion care in the U.S. following the Roe v. Wade repeal, or the 2017 repeal of F.C.C. rules that limit internet providers from selling customers' online information. Some of these newspaper articles contain recommendations from security experts and describe the security behavior of people affected by these current events. Less renowned news venues, such as Fox News or the German Bild, mainly mentioned Tor and VPNs in listicles of generic security tips or VPN-provider-sponsored articles. Almost all VPN providers in our list had blogs, wikis, or Q&A platforms explaining the difference between Tor and VPN, discussing the use of Tor over VPN, advertising their VPN servers' Tor feature, or instructing readers how to use a VPN to access the dark net. On Twitter, we could search for tweets in the last seven days. Consequently, we found Tor and VPN mentioned in the context of current events, in our case (July 2022), mainly the Roe v. Wade repeal and safe access to abortion health care. Tweets are limited to 280 characters, so many contained only instructions or a short list of recommended security and privacy tools. On Reddit and Stack Exchange, we discovered threads that discuss combining Tor and VPNs in various communities. On Stack Exchange, most threads were in security, privacy, or system-administration-specific communities, but we also found discussions in cryptocurrency, fictional writing, and law-focused communities. All communities that we identified on Stack Exchange were also active on Reddit. However, we also found active discussions in communities related to online drug shopping, piracy, and conspiracy theories. We also found discussions about Tor and VPN use in location-specific, political, or religion-affiliated communities in the context of current events, such as protests or ongoing repression. On Reddit, we found some instances of (former) members of oppressive social communities discussing recommended safety and security precautions for leaving the community. We reference quotes from our documents in the format "<Venue><Document-ID>:<Quote-ID>" to convey the general source to readers while enabling us to trace the quote's origin in our corpus.

Behavioral beliefs. Behavioral beliefs are about the negative or positive consequences users might experience when following a specific behavior [22]. Information that informs these beliefs about Tor over VPN needs to cover the potential effects of the practice. Assessing the quality of security measures relies on threat models, i.e., a specific potential threat that a measure mitigates. Most documents in our corpus, especially news articles, and Twitter, contained only implicit, vague, or all-encompassing threat models, e.g., *"[w]hen you surf the internet, everyone is watching"* (News986:1).

VPN providers focused on the supposed dangers of Tor (*"There's a chance that your computer may be used as an end-relay for illegal activity. "* (VPN43:4)) and hiding activities from internet providers and the government. They often warned about the threat of untrustworthy VPNs (not themselves). VPN providers' articles also contradict other VPN providers or even their own statements about

the effects of Tor over VPN. These contradictions usually depended on the effects of Tor over VPN and VPN over Tor they discussed. Sometimes they just blended the effects of both approaches into one. Unsurprisingly, most VPN providers who covered the topic recommended using a VPN to access the Tor network and advertised their VPN servers' Tor feature if available. Discussions about the effects of Tor over VPN were far more polarizing on Reddit and Stack Exchange. The range of opinions covered: absolutely necessary, beneficial, useless but harmless, and outright dangerous. On the more Tor-specific Reddit and Stack Exchange communities, the most common repeated answer to these discussions was that *"[t]he need for one, or both, depends on the user threat model"* (Reddit724:8) and is not needed for most users. However, they usually do not explicitly discuss which personal threat models warrant using Tor over VPN. While some threat models are relevant for all internet users, e.g., credential stuffing with leaked password data or malware attacks on users' outdated software, many threats do not affect internet users equally. Tor over VPN as a security practice appears to belong in the latter category. In general, interested readers looking for an authoritative answer to the question of what effect Tor over VPN has need to read discussions and choose an answer they like.

Normative beliefs. Injunctive normative beliefs concern others' assumed approval or disapproval of certain behaviors, which creates a social influence on behavior [22]. Descriptive normative beliefs concern peoples' perceptions of typically performed behavior based on observations. Related security research refers to this behavior as social proof [11–13]. Our analysis focused on these descriptive norms, i.e., what users would find in information sources when they look for social proof about the appropriate use of Tor. In many online discussions and fewer news articles, we found examples of people who combined Tor and VPNs and whom readers may perceive as experts. Their expertise expressed itself in often detailed technical questions that did not question the usefulness of Tor over VPN. Our teaser image in Figure 1 of a scene in the popular TV series Money Heist is a similar situation: Rio, a bona fide hacker, explains to the Professor that he accesses the dark net with a VPN and the Tor Browser. People who consume TV series may build up para-social relationships with the characters in them [28]. Hence, a scene like that, while not an explicit form of security advice, can create a perceived norm for people looking for social proof. Some news articles and VPN provider sites also expressed these perceived norms directly: For example, that people who are *"very serious about browsing anonymity"* (VPN43:2) or *"super security-focused"* (News913:2) can use Tor over VPN, and that Tor *"is popular among computer-savvy circles"* (News913:2). Consequently, some discussion participants assumed that using Tor over VPN is already a general practice, e.g., *"Pretty sure most Tor users do [use both at the same time]"* (Reddit708:6) or that using Tor by itself goes against best practice.

Perceived norms also manifest as jokes and memes about layering multiple security mechanisms on top of each other, e.g., *"OP is behind 3 VPNs, 2 reddit acounts [sic], a proxy, and also using TOR."* (Reddit708:2) or *"We can't find him! He's using 7 proxies in incognito mode!!"* (Reddit661:2). These comments seem to refer to the "Good Luck, I'm Behind 7 Proxies" meme, shown in Figure 5, which originated as a sarcastic retort on 4chan in 2007 [43].

While we sometimes had the impression that these references were sarcastic, discussion participants seemed to take them seriously more often than not, illustrated by this comment on Reddit: *"Seems the safest path is You -> Tor -> VPN -> Tor -> website."* (Reddit661:9); or an answer on Stack Exchange discussing the following setup: *"You –> VPN Provider (Probably listening by NSA) –> Tor –> Proxy (Any free proxy) –> VPN Provider(Another one listening by NSA) –> Tor –> Proxy (Any free proxy) –> Target"* (SE96:4).

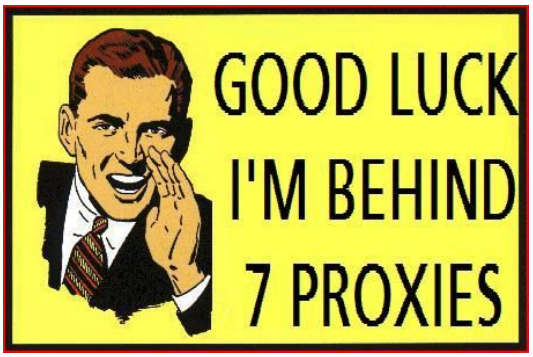


Fig. 5. The "Good Luck, I'm Behind 7 Proxies" meme, a sarcastic catchphrase that originated on 4chan in 2007 [43]

Appeals to authority on the subject also influence perceived norms. Discussion participants on Reddit and Stack Exchange regularly point to Tor- or Tails[2]-affiliated sites to strengthen their arguments. To a similar effect, news articles also include quotes and recommendations from renowned security experts. Other factors also establish a norm of combining Tor and VPNs. People often recommend using Tor over VPN without justifying it any further or explaining its effects: *"Tor + VPN is a good combo"* (Reddit708:5). Several VPN providers officially support Tor so that customers do not require the Tor Browser to access onion links, which lends credibility to the idea of combining Tor with a VPN. Thus, we found several indicators that information sources influence normative beliefs and offer social proof for using Tor over VPN.

Control beliefs. Control beliefs focus on personal and environmental aspects that support or hinder behavior [22]. Consequently, these beliefs impact people's perceived self-efficacy. Our corpus contained little information that could affect people's self-efficacy. We found specific step-by-step instructions on combining Tor and VPNs to achieve a particular goal, such as accessing the dark web safely or buying drugs online. These instructions usually did not explain the purpose or security of that practice. Achieving the primary goal is the only relevant aspect of the instructions. Other sources emphasize that using Tor over VPN or even the VPN-supported Tor feature is easy. That so few information sources cover these aspects could mean that VPNs and the Tor Browser are already easy to use or that the target demographic already has a high level of self-efficacy.

Misinterpretation of security advice involving Tor and VPNs. During the filtering process, we unexpectedly found that some sources contain ambiguous information on the combination of Tor and VPN, allowing for misinterpretation. Misinterpreting information sources may also impact users' normative beliefs about recommended practices. We included these ambiguous information sources in our corpus and labeled them with separate codes. We found two types of these information sources: First, we found lists of recommended security practices. While they do not directly recommend a combination of Tor and VPNs, these lists usually imply that users can combine all recommendations on that list and that each security advice provides additional security. We found these types of lists in news articles and social media posts. Second, we found ambiguous phrasing that may imply a combination of Tor and VPN, even when the author may have meant a separate use of these technologies. For example, a news article mentioned that internet users could use *"security tools like VPNs and Tor browser to bypass censorship"* (News970:1) and a Reddit

[2]Tails is a security-focused Linux distribution. It routes all data traffic through the Tor network.

user advised others to *"learn to how to use a VPN and Tor"* (Reddit435:2). This issue also affects statements from established security experts in well-known newspapers such as The New York Times, which said that *"many Russians were skilled at using it [Tor] and VPNs [for evading censorship]"* (News932:2). Journalists often look for short quotes from outside sources, making it difficult for experts to convey detailed information and recommendations. We found an example that illustrates how users misinterpret ambiguous information sources: In tweet Twitter1000:8, someone responded to the U.S. politician Alexandria Ocasio-Cortez, who disseminated a summarized version of EFF's digital safety tips for people seeking an abortion and providers of abortion support. The original message recommended *"steps like using a VPN, throwaway email addresses, private browsers"* and the commentator chimed in to say *"Yeah people get a tor and vpn and use duck duck go"*.

Misconceptions of the technology behind Tor and VPNs. User perception of the technology behind Tor and VPN may impact users' behavioral beliefs about Tor over VPN. Some people in online discussions saw the Tor browser as just one possible choice of privacy-preserving browser, listing it at the same level as Brave and the DuckDuckGo Privacy Browser, e.g., the Twitter user who recommended to *"Get a VPN, download tor or DuckDuckGo [Privacy Browser]"* (Twitter1000:33). For users who think of the Tor browser as just a browser with some extra features, it may seem intuitive that a VPN is required to protect the traffic. A group of people views Tor as just a particular type of VPN, applying the same threat models to both. They may have been familiar with VPNs before learning about Tor and inadvertently transferred their conception from one technology to another. Story et al. [62] found that experience with VPNs and the Tor browser is associated with confusion about these tools' protection. Lastly, we also identified many arguments centering around the concept of layering security mechanisms, where discussion participants implicitly applied the Swiss cheese model [55] to minimize risks. While this is not a misconception, it is a specific conception of security tools related to users' control beliefs. Applying this concept of "cumulative act effect" too broadly may lead to the conclusion that it is always helpful to add additional security tools.

7 DISCUSSION

7.1 Users Practice Tor over VPN with the Intention to Improve their General Security and Anonymity

We measured that 6.23% of connections to the Tor network originate from VPNs (see Section 4), corresponding to an estimated 140,000 daily users[3] at the time of writing. As our survey (see Section 5) indicates, most do not accidentally practice Tor over VPN. They intend to improve the security and start their VPN client specifically for their Tor session. They believe it increases their general security and protects them from unspecified dangers on the Tor network. Our document analysis of background sources (see Section 6) suggests that this can be explained by VPN providers who overemphasize or misrepresent the potential dangers of Tor in order to promote their products. We also found many recommendations to use Tor over VPN without describing the practice's purpose and effects, which explains why users expect general security benefits. This belief in security benefits is an issue since users expect more security and anonymity when not warranted. In the case of risk compensation behavior [60], this may result in net-negative security. Additionally, users have a limited compliance budget [5] for security practices. Hence, security practices with inconclusive benefits should be replaced by ones that provide general security benefits, such as using a password manager to create secure passwords or making backups to mitigate data loss.

[3] https://metrics.torproject.org/userstats-relay-country.html

7.2 Security Folklore Appears to Spread through Normative Beliefs

Results from our user survey (in Section 5) and document analysis (in Section 6) suggest that normative beliefs based on social proof contribute most to spreading security folklore. While we could find detailed discussions on the purpose and effects of Tor over VPN (behavioral beliefs) in expert communities on Reddit and Stack Exchange, we rarely found them in other places. These discussions included contradicting opinions and often had no clear outcome. However, VPN providers' websites gave some suggestions on Tor over VPN's purpose, i.e., hiding Tor use from internet providers and protecting from the supposed dangers of Tor, which also came up again in the expert discussions on Reddit and Stack Exchange as well as our survey on users' intentions and beliefs.

In contrast, we found evidence for normative beliefs and quotes that provide social proof in all types of documents. Quotes that support normative beliefs come without threat models and descriptions of consequences, focusing instead on the people that implement the practice. Discussions that assume Tor over VPN is a general practice, news articles that say that "security-focused users" implement it, and pop-culture media mentions of the practice all contribute to establishing perceived norms of security practice. Even the security community is guilty of doing this: "Use Signal, use Tor" is a really common but criticized form of security advice without specific threat models. Even Edward Snowden tweeting the advice in 2016 [19] has firmly established it as a perceived norm. Today, memes, stickers, and t-shirts include this security advice. Hassoun et al. [32] found that Gen Zers' informational and social needs are inseparably entangled, using information to orient themselves socially and define their emerging identities. A similar entanglement of informational and social needs could explain some of our results on normative beliefs.

Hence, we argue that normative beliefs contribute to the spread of security folklore for social reasons when people look to others to decide which behavior is appropriately secure. Of course, this works better when recommended practices are easy to implement, i.e., when self-efficacy is not an issue (control beliefs). Lastly, the role of normative beliefs in spreading this Tor over VPN use is also underlined by the fact that most participants in the survey expected general security benefits, i.e., they did not associate any behavioral beliefs with the practice.

In Section 3, we described two types of normative beliefs: injunctive norms that focus on others' potential judgment of one's behavior and descriptive norms that stem from observing others' behavior. In general, injunctive norms have more influence than descriptive norms regarding actually swaying behaviors. While injunctive and descriptive normative beliefs likely play a role in the analyzed information sources, we deliberately focused on descriptive norms. Researching injunctive norms in an online setting is difficult since it is unclear how much discussion participants care about their peers' judgment – they never really have to reveal their actual behavior to them.

While the results of this work demonstrate which type of beliefs affect the spread of this particular security behavior, it raises new questions for future work about the exact origins of advice and beliefs and their spreading path across platforms.

7.3 Giving Effective Security Advice for Specific Use-Cases

Our case study illustrates some essential points for providing effective security advice: First, demonstrated security behavior or stories of security practices influence normative beliefs and should be treated as a form of (unintentional) security advice. We should try to communicate the minimal set of security advice with the most practicality and general security benefits [58]. However, it is not possible to only tell stories about security practices that have general security benefits. Hence, not only security advice [8, 26] but also stories about security practices, e.g., in news articles, need to include an explanation of their purpose.

Second, it is easy for readers to misinterpret security advice or stories about security practices, i.e., listicles that imply combination behavior, ambiguous phrasing, or lack of threat models. Therefore, checking for possible misinterpretations and providing extra space when necessary should be a required part of a responsible publication process on security topics.

Third, security is usually not the users' primary goal. Giving clear instructions on how to achieve the primary goal securely may be an effective method of giving security advice for exceptional use cases. VPN providers already apply this approach; many have articles instructing users to connect to a VPN before accessing the dark net.

Lastly, layering multiple security mechanisms seems to create a perceived sense of security, i.e., a secure experience. Avoiding these unwarranted secure experiences requires design approaches that take them into account. Designing the Tor browser differently so that it informs users who connect with a VPN of its security implications may combat an unwarranted perceived sense of security.

7.4 User Perceptions of Security Technology Shape their Behavior

In the user survey (Section 5) and the document analysis (Section 6), we found several cases where users' perceptions of the technology behind Tor and VPNs shaped their beliefs and self-reported behavior. We found that some users think of the Tor browser as a regular privacy-preserving browser that merely adds countermeasures at the application layer, e.g., protecting from behavioral online tracking by blocking third-party cookies. Such a belief would explain why these users think that protecting the network layer, i.e., the browsing traffic, with a VPN is beneficial for using the Tor browser securely – even though it is not.

Other users thought of Tor in the same way as VPNs and thus treated both in the same way, which confirms Story et al.'s [62] finding that users had trouble differentiating Tor and VPNs. We also found indications that users apply the Swiss cheese model of risk management [55] to justify layering security practices. Lastly, the responses to our user survey suggest that VPN providers have successfully created the perception that Tor is unsafe to use without a VPN. A part of the reason for these user perceptions and coping strategies is that the security technology behind Tor and VPNs and its effects are hard to communicate to laypeople.

While not all misconceptions are fixable, some general approaches tackle the misconception that the Tor browser is just a regular browser: names we use for software or features, visualizations that communicate inner workings, and user interaction that support users in building mental models. First, the name of the Tor software would ideally hint at its effect on the network layer while also making it more difficult to confuse with other types of privacy-preserving browsers. The potential effect on user perception did not seem to have been an important factor when renaming the "Tor Browser Bundle" to simply "Tor browser" nine years ago.[4] However, while a better name might help new users, it also has the potential to confuse the existing user base. Hence, future empirical user research must evaluate this option with care. Second, upon starting up the Tor browser, visualizing the ongoing connection attempts that are usually in the background may improve user understanding of the entire system. Third, upon detecting VPN use, user interaction in the form of prompts could explain that using a VPN in combination is potentially unnecessary. Nudging users in this way to question their beliefs and possibly change their behavior is a well-studied form of soft paternalism that helps users understand and choose appropriate options [3].

8 CONCLUSION

To understand how security folklore spreads and how it could facilitate useful security advice, we investigated a data-rich example, i.e., the practice of using Tor over VPN. Our case study consisted

[4] https://gitlab.torproject.org/legacy/trac/-/issues/11193

of three study parts aligned with the theory of reasoned action: We quantified behavior on real-world Tor traffic, surveyed users' intentions and beliefs that explain the behavior, and analyzed background information sources that affect user beliefs.

We found that the spread of security folklore relies on the normative beliefs of users looking for social proof [11–13] on how to use Tor securely. This effect appears especially important in cases such as Tor over VPN when threat models and consequences of security practices are hard to determine for laypeople. Thus, the security community may want to investigate how to effectively establish normative beliefs to spread useful security advice with generally applicable security benefits. However, to avoid the unintentional spread of security misinformation as normative beliefs, reports about security practices in newspaper articles and pop-culture media must explain their effects and accompanying threat models.

We also found that misinterpretations and misconceptions support erroneous beliefs about the security technology behind Tor and VPN and its effects.

Overall, our Tor-specific and generalizable findings suggest that careful phrasing is crucial for communicating information about security practices, and redesigning tools' secure experience through visualizations or UI friction may combat some misconceptions.

9 ACKNOWLEDGMENTS

We thank all reviewers for their feedback on improving our paper. The first two authors, Matthias Fassl and Alexander Ponticello, worked on this research as part of the Saarbrücken Graduate School of Computer Science, Saarland University.

REFERENCES

[1] Ruba Abu-Salma and Benjamin Livshits. 2020. Evaluating the End-User Experience of Private Browsing Mode. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–12. https://doi.org/10.1145/3313831.3376440

[2] Ruba Abu-Salma, M. Angela Sasse, Joseph Bonneau, Anastasia Danilova, Alena Naiakshina, and Matthew Smith. 2017. Obstacles to the Adoption of Secure Communication Tools. In *2017 IEEE Symposium on Security and Privacy (SP)*. IEEE, San Jose, CA, USA, 137–153. https://doi.org/10.1109/SP.2017.65

[3] Alessandro Acquisti, Idris Adjerid, Rebecca Balebako, Laura Brandimarte, Lorrie Faith Cranor, Saranga Komanduri, Pedro Giovanni Leon, Norman Sadeh, Florian Schaub, Manya Sleeper, Yang Wang, and Shomir Wilson. 2017. Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online. *Comput. Surveys* 50, 3 (Oct. 2017), 1–41. https://doi.org/10.1145/3054926

[4] Omer Akgul, Richard Roberts, Moses Namara, Dave Levin, and Michelle L. Mazurek. 2022. Investigating Influencer VPN Ads on YouTube. In *2022 IEEE Symposium on Security and Privacy (SP)*. IEEE, San Francisco, CA, USA, 876–892. https://doi.org/10.1109/SP46214.2022.9833633

[5] Adam Beautement, M. Angela Sasse, and Mike Wonham. 2008. The Compliance Budget: Managing Security Behaviour in Organisations. In *Proceedings of the 2008 New Security Paradigms Workshop (NSPW '08)*. ACM, Lake Tahoe, CA, USA, 47—58. https://doi.org/10.1145/1595676.1595684

[6] Veroniek Binkhorst, Tobias Fiebig, Katharina Krombholz, Wolter Pieters, and Katsiaryna Labunets. 2022. Security at the End of the Tunnel: The Anatomy of VPN Mental Models Among Experts and Non-Experts in a Corporate Context. In *31st USENIX Security Symposium (USENIX Security 22)*. USENIX Association, Boston, MA, USA, 3433–3450. https://www.usenix.org/conference/usenixsecurity22/presentation/binkhorst

[7] Alex Biryukov and Ivan Pustogarov. 2015. Bitcoin over Tor Isn't a Good Idea. In *2015 IEEE Symposium on Security and Privacy (SP)*. IEEE, San Jose, CA, 122–134. https://doi.org/10.1109/SP.2015.15

[8] Maia J. Boyd, Jamar L. Sullivan Jr., Marshini Chetty, and Blase Ur. 2021. Understanding the Security and Privacy Advice Given to Black Lives Matter Protesters. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama, Japan, 1–18. https://doi.org/10.1145/3411764.3445061

[9] Virginia Braun and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3, 2 (Jan. 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[10] Jan Harold Brunvand. 1978. *The Study of American Folklore: An Introduction* (second ed.). Norton, New York.

[11] Robert B. Cialdini. 2009. *Influence: Science and Practice*. Pearson Education.

[12] Sauvik Das, Tiffany Hyun-Jin Kim, Laura A Dabbish, and Jason I Hong. 2014. The Effect of Social Influence on Security Sensitivity. In *10th Symposium On Usable Privacy and Security (SOUPS)*. USENIX Association, 143—157.

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

[14] Albese Demjaha, Jonathan Spring, Ingolf Becker, Simon Parkin, and Angela Sasse. 2018. Metaphors Considered Harmful? An Exploratory Study of the Effectiveness of Functional Metaphors for End-to-End Encryption. In *Proceedings of the Workshop on Usable Security (USEC)*. Internet Society, San Diego, CA, USA. https://doi.org/10.14722/usec.2018.23015

[15] Roger Dingledine. 2012. [Tor-Talk] Tor plus VPN (Was Re: Hi All!). Retrieved 2022-11-23 from https://lists.torproject.org/pipermail/tor-talk/2012-January/022917.html

[16] Roger Dingledine, Nicholas Hopper, George Kadianakis, and Nick Mathewson. 2014. One Fast Guard for Life (or 9 Months). In *Proceedings of the 14th Privacy Enhancing Technologies (PETS)*. Amsterdam, Netherlands. https://doi.org/10.1.1.645.7692

[17] Roger Dingledine and Nick Mathewson. 2021. Tor Path Specification. Retrieved 2022-11-22 from https://github.com/torproject/torspec/blob/master/path-spec.txt

[18] Agnieszka Dutkowska-Zuk, Austin Hounsel, Andre Xiong, Molly Roberts, Brandon Stewart, Marshini Chetty, and Nick Feamster. 2020. Understanding How and Why University Students Use Virtual Private Networks. arXiv. https://doi.org/10.48550/ARXIV.2002.11834 arXiv:2002.11834

[19] Edward Snowden [@Snowden]. 2016. Use Tor. Use Signal. https://t.co/VLvBsbVHKs. Retrieved 2022-11-22 from https://twitter.com/Snowden/status/778592275144314884

[20] Ellis Fenske, Akshaya Mani, Aaron Johnson, and Micah Sherr. 2017. Distributed Measurement with Private Set-Union Cardinality. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS)*. ACM, Dallas, TX, USA, 2295—2312. https://doi.org/10.1145/3133956.3134034

[21] Casey Fiesler and Nicholas Proferes. 2018. "Participant" Perceptions of Twitter Research Ethics. *Social Media + Society* 4, 1 (Jan. 2018), 205630511876336. https://doi.org/10.1177/2056305118763366

[22] Martin Fishbein and Icek Ajzen. 2010. *Predicting and Changing Behavior: The Reasoned Action Approach*. Psychology Press, New York.

[23] George Forman. 2005. Counting Positives Accurately Despite Inaccurate Classification. In *16th European Conference on Machine Learning (ECML)*, João Gama, Rui Camacho, Pavel B. Brazdil, Alípio Mário Jorge, and Luís Torgo (Eds.). Springer Berlin Heidelberg, Porto, Portugal, 564–575. https://doi.org/10.1007/11564096_55

[24] Kevin Gallagher, Sameer Patil, and Nasir Memon. 2017. New Me: Understanding Expert and Non-Expert Perceptions and Usage of the Tor Anonymity Network. In *Proceedings of the Thirteenth Symposium on Usable Privacy and Security (SOUPS)*. USENIX Association, Santa Clara, CA, USA, 385–398. https://www.usenix.org/system/files/conference/soups2017/soups2017-gallagher.pdf

[25] Xianyi Gao, Yulong Yang, Huiqing Fu, Janne Lindqvist, and Yang Wang. 2014. Private Browsing: An Inquiry on Usability and Privacy Protection. In *Proceedings of the 13th Workshop on Privacy in the Electronic Society*. ACM, Scottsdale Arizona USA, 97–106. https://doi.org/10.1145/2665943.2665953

[26] Christine Geeng, Mike Harris, Elissa Redmiles, and Franziska Roesner. 2022. "Like Lesbians Walking the Perimeter": Experiences of U.S. LGBTQ+ Folks With Online Security, Safety, and Privacy Advice. In *31st USENIX Security Symposium (USENIX Security 22)*. USENIX Association, Boston, MA, USA, 305–322. https://www.usenix.org/conference/usenixsecurity22/presentation/geeng

[27] Sarah Gilbert, Katie Shilton, and Jessica Vitak. 2023. When Research Is the Context: Cross-platform User Expectations for Social Media Data Reuse. *Big Data & Society* 10, 1 (Jan. 2023), 205395172311641. https://doi.org/10.1177/20539517231164108

[28] David C. Giles. 2002. Parasocial Interaction: A Review of the Literature and a Model for Future Research. *Media Psychology* 4, 3 (Aug. 2002), 279–305. https://doi.org/10.1207/S1532785XMEP0403_04

[29] David J. Gunkel. 2018. The Relational Turn: Third Wave HCI and Phenomenology. In *New Directions in Third Wave Human-Computer Interaction: Volume 1 - Technologies*, Michael Filimowicz and Veronika Tzankova (Eds.). Springer International Publishing, Cham, 11–24. https://doi.org/10.1007/978-3-319-73356-2_2

[30] Hana Habib, Jessica Colnago, Vidya Gopalakrishnan, Sarah Pearman, Jeremy Thomas, Alessandro Acquisti, Nicolas Christin, and Lorrie Faith Cranor. 2018. Away From Prying Eyes: Analyzing Usage and Understanding of Private Browsing. In *Proceedings of the Fourteenth Symposium on Usable Privacy and Security (SOUPS)*. USENIX Association, Baltimore, MD, USA, 159–175. https://www.usenix.org/system/files/conference/soups2018/soups2018-habib-prying.pdf

[31] David Harborth, Sebastian Pape, and Kai Rannenberg. 2020. Explaining the Technology Use Behavior of Privacy-Enhancing Technologies: The Case of Tor and JonDonym. *Proceedings on Privacy Enhancing Technologies* 2020, 2 (April 2020), 111–128. https://doi.org/10.2478/popets-2020-0020

[32] Amelia Hassoun, Ian Beacock, Sunny Consolvo, Beth Goldberg, Patrick Gage Kelley, and Daniel M. Russell. 2023. Practicing Information Sensibility: How Gen Z Engages with Online Information. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 662, 17 pages. https://doi.org/10.1145/3544548.3581328

[33] Muhammad Ikram, Narseo Vallina-Rodriguez, Suranga Seneviratne, Mohamed Ali Kaafar, and Vern Paxson. 2016. An Analysis of the Privacy and Security Risks of Android VPN Permission-enabled Apps. In *Proceedings of the 2016 Internet Measurement Conference*. ACM, Santa Monica, CA, USA, 349–364. https://doi.org/10.1145/2987443.2987471

[34] Rob Jansen and Aaron Johnson. 2016. Safely Measuring Tor. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS '16)*. ACM, Vienna, Austria, 1553–1567. https://doi.org/10.1145/2976749.2978310

[35] Ruogu Kang, Stephanie Brown, Laura Dabbish, and Sara Kiesler. 2014. Privacy Attitudes of Mechanical Turk Workers and the U.S. Public. In *Proceedings of the Tenth Symposium On Usable Privacy and Security (SOUPS)*. USENIX Association, Menlo Park, CA, USA, 37–49. https://doi.org/10.5555/3235838.3235842

[36] Ruogu Kang, Laura Dabbish, Nathaniel Fruchter, and Sara Kiesler. 2015. "My Data Just Goes Everywhere": User Mental Models of the Internet and Implications for Privacy and Security. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015) (SOUPS)*. USENIX Association, Ottawa, ON, Canada, 39–52. https://www.usenix.org/system/files/conference/soups2015/soups15-paper-kang.pdf

[37] Mohammad Taha Khan, Joe DeBlasio, Geoffrey M. Voelker, Alex C. Snoeren, Chris Kanich, and Narseo Vallina-Rodriguez. 2018. An Empirical Analysis of the Commercial VPN Ecosystem. In *Proceedings of the Internet Measurement Conference 2018 (IMC '18)*. ACM, Boston, MA, USA, 443–456. https://doi.org/10.1145/3278532.3278570

[38] Alexandra Mai, Katharina Pfeffer, Matthias Gusenbauer, Edgar Weippl, and Katharina Krombholz. 2020. User Mental Models of Cryptocurrency Systems - A Grounded Theory Approach. In *Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020)*. USENIX Association, 341–358. https://www.usenix.org/conference/soups2020/presentation/mai

[39] Akshaya Mani, T. Wilson-Brown, Rob Jansen, Aaron Johnson, and Micah Sherr. 2018. Understanding Tor Usage with Privacy-Preserving Measurement. In *Proceedings of the Internet Measurement Conference 2018 (IMC)*. ACM, Boston, MA, USA, 175–187. https://doi.org/10.1145/3278532.3278549

[40] Mara Mather and Marcia K. Johnson. 2000. Choice-Supportive Source Monitoring: Do Our Decisions Seem Better to Us as We Age? *Psychology and Aging* 15, 4 (2000), 596–606. https://doi.org/10.1037/0882-7974.15.4.596

[41] Niels Raabjerg Mathiasen and Susanne Bødker. 2008. Threats or Threads: From Usable Security to Secure Experience?. In *Proceedings of the 5th Nordic Conference on Human-Computer Interaction: Building Bridges (NordiCHI)*. ACM, Lund, Sweden, 283–289. https://doi.org/10.1145/1463160.1463191

[42] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and Inter-rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 1–23. https://doi.org/10.1145/3359174

[43] Know Your Meme. 2010. Good Luck, I'm Behind 7 Proxies. Retrieved 2022-11-22 from https://knowyourmeme.com/memes/good-luck-im-behind-7-proxies

[44] Rebecca Moody. 2020. VPN Market Report 2022: Who's Got the Biggest VPN Market Share? Retrieved 2022-11-22 from https://www.comparitech.com/blog/vpn-privacy/vpn-market-share-report/

[45] Moses Namara, Daricia Wilkinson, Kelly Caine, and Bart P. Knijnenburg. 2020. Emotional and Practical Considerations Towards the Adoption and Abandonment of VPNs as a Privacy-Enhancing Technology. *Proceedings on Privacy Enhancing Technologies* 2020, 1 (2020), 83—102. https://doi.org/10.2478/popets-2020-0006

[46] Nic Newman, Richard Fletcher, Craig T Robertson, Kirsten Eddy, and Rasmus Kleis Nielsen. 2022. *Reuters Institute Digital News Report 2022*. Technical Report. 164 pages. Retrieved 2022-11-22 from https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2022-06/Digital_News-Report_2022.pdf

[47] nusenu. 2021. OrNetStats. Retrieved 2021-07-05 from https://nusenu.github.io/OrNetStats/

[48] Anna-Marie Ortloff, Matthias Fassl, Alexander Ponticello, Florin Martius, Anne Mertens, Katharina Krombholz, and Matthew Smith. 2023. Different Researchers, Different Results? Analyzing the Influence of Researcher Experience and Data Type During Qualitative Analysis of an Interview and Survey Study on Security Advice. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–21. https://doi.org/10.1145/3544548.3580766

[49] Katharina Pfeffer, Alexandra Mai, Edgar Weippl, Emilee Rader, and Katharina Krombholz. 2022. Replication: Stories as Informal Lessons about Security. In *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*. USENIX Association, Boston, MA, 1–18. https://www.usenix.org/conference/soups2022/presentation/pfeffer

[50] Álex Pina, Esther Martínez Lobato, Javier Gómez Santander, Pablo Roa, Fernando Sancristóbal, Esther Morales, David Barrocal (Writers), and Javier Quintas (Director). 2017. Money Heist - A Matter of Efficiency (Season 2, Part 3). https://www.imdb.com/title/tt6851508/

[51] Emilee Rader and Janine Slaker. 2017. The Importance of Visibility for Folk Theories of Sensor Data. In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*. USENIX Association, Santa Clara, CA, 257–270. https://www.usenix.org/conference/soups2017/technical-sessions/presentation/rader

[52] Emilee Rader, Rick Wash, and Brandon Brooks. 2012. Stories as Informal Lessons about Security. In *Proceedings of the Eighth Symposium on Usable Privacy and Security (SOUPS '12)*. ACM, Washington D.C., USA. https://doi.org/10.1145/2335356.2335364

[53] Reethika Ramesh, Leonid Evdokimov, Diwen Xue, and Roya Ensafi. 2022. VPNalyzer: Systematic Investigation of the VPN Ecosystem. In *Proceedings 2022 Network and Distributed System Security Symposium*. Internet Society, San Diego, CA, USA. https://doi.org/10.14722/ndss.2022.24285

[54] Reethika Ramesh, Anjali Vyas, and Roya Ensafi. 2023. "All of Them Claim to Be the Best": Multi-perspective Study of VPN Users and VPN Providers. In *32nd USENIX Security Symposium (USENIX Security 23)*. USENIX Association, Anaheim, CA. https://www.usenix.org/conference/usenixsecurity23/presentation/ramesh

[55] James Reason. 1990. The Contribution of Latent Human Failures to the Breakdown of Complex Systems. *Philosophical Transactions of the Royal Society B* 327 (1990), 475–484. https://doi.org/10.1098/rstb.1990.0090

[56] Tor Reddit. 2021. FAQ: "Should I Use a VPN with Tor?". Retrieved 2022-11-22 from https://old.reddit.com/r/TOR/wiki/index#wiki_should_i_use_a_vpn_with_tor.3F_tor_over_vpn.2C_or_vpn_over_tor.3F

[57] Elissa M Redmiles, Sean Kross, and Michelle L Mazurek. 2016. How I Learned to Be Secure: A Census-Representative Survey of Security Advice Sources and Behavior. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS '16)*. ACM, Vienna, Austria, 666—677. https://doi.org/10.1145/2976749.2978307

[58] Elissa M. Redmiles, Noel Warford, Amritha Jayanti, and Aravind Koneru. 2020. A Comprehensive Quality Evaluation of Security and Privacy Advice on the Web. In *Proceedings of the 29th USENIX Security Symposium (USENIX Security 2020)*. USENIX Association, 89–108. https://www.usenix.org/conference/usenixsecurity20/presentation/redmiles

[59] Robert W. Reeder, Iulia Ion, and Sunny Consolvo. 2017. 152 Simple Steps to Stay Safe Online: Security Advice for Non-Tech-Savvy Users. *IEEE Security & Privacy* 15, 5 (2017), 55—64. https://doi.org/10.1109/msp.2017.3681050

[60] Karen Renaud and Merrill Warkentin. 2017. Risk Homeostasis in Information Security: Challenges in Confirming Existence and Verifying Impact. In *Proceedings of the 2017 New Security Paradigms Workshop (NSPW 2017)*. ACM, Santa Cruz, CA, USA, 57—69. https://doi.org/10.1145/3171533.3171534

[61] Bruce Schneier. 2008. The Psychology of Security. In *AFRICACRYPT*. 30. https://doi.org/10.1007/978-3-540-68164-9_5

[62] Peter Story, Daniel Smullen, Yaxing Yao, Alessandro Acquisti, Lorrie Faith Cranor, Norman Sadeh, and Florian Schaub. 2021. Awareness, Adoption, and Misconceptions of Web Privacy Tools. *Proceedings on Privacy Enhancing Technologies* 2021, 3 (2021), 308—333. https://doi.org/10.2478/popets-2021-0049

[63] Jenny Tang, Eleanor Birrell, and Ada Lerner. 2022. Replication: How Well Do My Results Generalize Now? The External Validity of Online Privacy and Security Surveys. In *Proceedings of the Eighteenth Symposium on Usable Privacy and Security (SOUPS)*. USENIX Association, Boston, MA, USA. https://www.usenix.org/conference/soups2022/presentation/tang

[64] TOP10VPN and globalwebindex. 2020. *Global VPN Usage Report 2020: An Exploration of VPNs and Their Users around the World*. Technical Report. 19 pages. Retrieved 2022-11-22 from https://www.top10vpn.com/assets/2020/03/Top10VPN-GWI-Global-VPN-Usage-Report-2020.pdf

[65] TorProject Trac. 2022. Tor + VPN. Retrieved 2022-11-22 from https://trac.torproject.org/projects/tor/wiki/doc/TorPlusVPN

[66] Matt Traudt. 2016. VPN + Tor: Not Necessarily a Net Gain. Retrieved 2022-11-22 from https://matt.traudt.xyz/posts/2016-11-12-vpn-tor-not-net-gain/

[67] Blase Ur, Fumiko Noma, Jonathan Bees, Sean M Segreti, Richard Shay, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2015. "I Added '!' At the End to Make It Secure": Observing Password Creation in the Lab. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*. USENIX Association, Ottawa, ON, Canada, 123–140. https://www.usenix.org/conference/soups2015/proceedings/presentation/ur

[68] Rick Wash. 2010. Folk Models of Home Computer Security. In *Proceedings of the Sixth Symposium on Usable Privacy and Security (SOUPS '10)*. ACM, Redmond, WA, USA, 1–16. https://doi.org/10.1145/1837110.1837125

[69] Zachary Weinberg, Shinyoung Cho, Nicolas Christin, Vyas Sekar, and Phillipa Gill. 2018. How to Catch When Proxies Lie: Verifying the Physical Locations of Network Proxies with Active Geolocation. In *Proceedings of the Internet Measurement Conference 2018*. ACM, Boston MA USA, 203–217. https://doi.org/10.1145/3278532.3278551

[70] Yixin Zou, Kevin Roundy, Acar Tamersoy, Saurabh Shintre, Johann Roturier, and Florian Schaub. 2020. Examining the Adoption and Abandonment of Security, Privacy, and Identity Theft Protection Practices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI)*. ACM, Honolulu, HI, USA, 1–15. https://doi.org/10.1145/3313831.3376570

Received January 2023; revised April 2023; accepted May 2023