Replication:"No one can hack my mind" - 10 years later: An update and outlook on experts' and non-experts' security practices and advice

Anna-Marie Ortloff
University of Bonn

Jenny Tang
Carnegie Mellon University

Arthi Arumugam
University of Bonn

Daniel Huschina
University of Bonn

Lisa Geierhaas
University of Bonn

Florin Martius
University of Bonn

Luisa Jansen
University of Bern

Kolja von der Twer
University of Bonn

Lilly Jungbluth
University of Bonn

Matthew Smith
University of Bonn, Fraunhofer FKIE

Abstract

In 2015, Ion, Reeder, and Consolvo studied IT security advice and self-reported security behavior of experts and non-experts. In 2019, Busse et al. replicated this study and found only minor changes in expert advice and non-expert behavior, with persisting differences between the two groups. Now, 10 years later, we replicated the study with an updated survey and compared our results to both prior studies. Additionally, we interviewed security experts and asked them for their views on the past and future of IT security advice. We report the current state of security behavior and advice based on two survey samples: one non-expert (N=990), and one expert sample (N=75) and an additional expert interview sample (N=35). We identified notable changes in reported security behavior for both experts and non-experts, including that experts and non-experts are beginning to adopt new security practices in authentication. The expert interviews show a path forward, with experts hoping for more improvements to usability and targeted advice for specific user and device-contexts.

1 Introduction

Staying secure online is still a challenge to users today as in the early days of Usable Privacy & Security (UPS) [3, 7, 64, 94]. Advice on staying secure has therefore been a much researched topic in our field [11, 15, 28, 37, 43, 57, 67–71]. In 2015, Ion et al. published a study comparing expert and non-expert security advice and behavior and found that experts and non-experts differ in their behavior, but also that experts do not strictly agree on a set of advice [46]. Their study was later replicated by Busse et al. and published in 2019 [17]. The replication showed that many of the challenges in the first study were still apparent 4 years later, but also suggested some development [17]. Research has shown that advice is still fragmented—making it difficult for non-expert users to prioritize between different advice, as well as actually putting advice into practice—as usability is still a hurdle to achieving security [69]. Now, almost ten years after the original "No one can hack my mind" study was published, we replicated the study again and took a look back at ten years of security advice and behavior by experts and non-experts. Specifically, we investigated the following research questions:

RQ1: How has security behavior and advice developed over the past ten years?

RQ2: How do experts and non-experts differ in their security behavior?

RQ3: What are security experts' perspectives on the development of security behavior in the past and in the future?

Our work identifies a larger shift in security behavior and advice than the first replication from 2019 [17]. Over the duration of ten years, some behaviors, e.g., using two-factor authentication (2FA), have seen a strong rise in popularity, while others, such as the use of antivirus, have become less relevant. This establishes a time frame for researchers to observe such developments. In addition to this main contribution, we provide an updated overview of current end-user security practices and discuss both challenges and successes for the UPS community. Password manager use is an example of an ongoing challenge, where expert advice diverges from end user practice. On the other hand, our study also revealed UPS success stories, e.g., increased adoption of 2FA. Looking ahead, experts express a hope for more user-friendly and context-specific security advice in the future, such as a shift in focus away from computers towards the now more prevalent phones or tablets. They also support future developments moving toward security measures that are built into systems by default, circumventing user error, e.g., automatic updates. The UPS

USENIX Symposium on Usable Privacy and Security (SOUPS) 2025.
August 10–12, 2025, Seattle, WA, United States.

community should monitor the changing status of security advice in these areas in longitudinally-oriented research. We support this goal by providing a replication package on OSF.

2 Related Work

We present prior work on security advice and provide an overview of the two studies we replicate in this work [17, 46]. We also present the development of key areas of security behavior and advice since Busse et al.'s iteration of this study.

2.1 Challenges in End-User Security Advice

IT security advice has been investigated by numerous studies [17, 28, 37, 43, 46, 67–69, 71, 91, 92]. Prior work found a gap between expert security advice and end-user behavior [17, 46]. One reason for this gap is that end users are overwhelmed by the volume of security recommendations available, making it difficult to discern which advice is essential and actionable [69]. Users may reject security advice for several reasons: they may perceive it as marketing-driven, as a threat to their privacy, or as coming from untrusted sources [68]. One challenge in security education is that users must trust that their security-related actions will yield positive outcomes, despite these benefits being difficult to see or verify [18, 28], leading to rejection of advice due to a poor cost-benefit trade-off [43]. Also, users may believe that such practices are unnecessary due to differing mental models of security [91].

Compounding this issue, even security experts do not agree on best practices, making it challenging to establish a clear and consistent set of recommendations [17, 46, 69]. Security beliefs vary across different demographic groups [92], highlighting that a one-size-fits-all approach to security education is ineffective. In fact, security recommendations often need to be tailored to specific user groups. For example, older adults have distinct security concerns and may require communication strategies that differ from those for younger users [56]. Similarly, software developers require security guidance that integrates seamlessly into their workflows [2]. Activists, who often operate under high-risk conditions, need advice that is contextually relevant to their security threats [15] and LGBTQ+ individuals may face unique privacy and security challenges that necessitate specialized guidance [37].

2.2 Replicated Studies

We base our work on two studies, the original "No one can hack my mind" study from 2014 by Ion et al. [46], published in 2015, and its replication by Busse et al., conducted in 2018, published in 2019 [17]. Each consisted of interviews with security experts in addition to surveys with security experts and non-experts on the topic of security behavior and advice. A comparison of study design and goals between these publications and our re-replication is in Table 1.

The original study by Ion et al. [46] highlighted that experts and non-experts approach online security differently: Experts rated some of the non-expert practices positively, but non-experts did not comply with other practices that received high ratings from experts, the most prominent of which being installing system updates, using a password manager and using two-factor authentication. They discussed more detailed results for four categories of advice: *software updates*, *antivirus software*, *password management*, and *mindfulness*. The replication by Busse et al. [17] showed that many of the topics relevant in the original work were largely unchanged in 2018. One development was that non-experts preferred to use browser extensions to block cookies rather than deleting them manually, compared to the original study. In their evaluation of the compound question asking for ratings of effectiveness and realism within a single question, they identified several areas where experts' effectiveness and realism ratings diverged: password security, two-factor authentication, links and attachments, and application updates. These areas emerged as areas for improvement for the UPS community.

2.3 Security Behavior

In the following, we describe areas that emerged from the original studies [17, 46] as important in terms of end user behavior and how the fields changed over time. Early research on security behavior, such as phishing detection, indicated that end users had difficulties distinguishing phishing sites from legitimate ones, often ignoring phishing indicators or security warnings [24, 27, 96]. The challenge has become even greater as attackers increasingly use HTTPS, making traditional security cues less reliable [49, 51, 84]. Now, new attack vectors, such as voice phishing and SMS phishing, have become more prevalent, particularly with the rise of social media-based scams [6]. To counter these threats, machine learning techniques have been applied for phishing detection [72, 78, 85, 87, 90]. However, adversaries have also exploited these techniques to evade detection, shifting the responsibility for identifying phishing attacks back onto users [5, 61, 76].

Beyond phishing, general system security is also influenced by how users handle software updates. Keeping software up to date is one of the most effective ways to mitigate security vulnerabilities [93]. However, research shows that many users fail to associate updates with security benefits, leading to delays or avoidance [28, 32, 54, 67, 89]. To reduce user burden, automatic and silent updates have been introduced, making security maintenance more seamless [25, 30, 89]. This evolution might change advice about software updates to be seen as less important as it once was.

Another domain where traditional security practices have evolved is malware protection. Previously, third-party antivirus (AV) software was widely recommended to protect against viruses and malware. However, modern operating systems have integrated built-in security features, reducing the

		Ion et al. [46]	Busse et al. [17]	This paper
Expert Interviews	Recruitment	BlackHat, DefCon, USENIX Security	CeBIT international trade fair	BlackHat London, C3, USP event
	Security experience for inclusion	5+ years	not specified	1+ years
	Sample *N*	40	40	35
	Study goal	develop survey	evaluate survey design	update survey questions
Expert Survey	Recruitment	Google Security Blog, personal contacts	social media, personal contacts	social media, personal contacts, Freelancer.com
	Security experience for inclusion	5+ years	1+ years	1+ years
	Survey design	compound	both compound and separate (A/B design)	compound; added questions on new behaviors
Non-Expert Survey	Recruitment	Mturk	Mturk	Prolific
	Target demographic	US	US	US representative
	Survey design	separate	separate	separate; added questions on new behaviors

Table 1: Methodological comparison for all studies by Ion et al. [46], by Busse et al. [17], and in this paper. Compound = compound questions used for advice rating; separate = separate questions used for rating effectiveness and realism of advice

necessity for separate AV solutions [35, 95]. For instance, Windows 11 offers a trusted boot process, built-in encryption, and network security, alongside native virus and threat protection [52]. Similarly, macOS integrates hardware-based protections and advanced encryption to enhance security [8].

Passwords have been used for authentication since the 1960s [13] and remain prevalent for end users. However, challenges such as weak passwords and reuse across multiple sites make them susceptible to attacks [16, 22, 31]. To address usability and security concerns, password managers have been introduced, allowing users to generate and store unique, strong passwords across multiple accounts, significantly reducing the risks of credential re-use [36, 50, 60]. Additionally, two-factor authentication was implemented more and more to secure user accounts [12]. Modern authentication has evolved further, with growing adoption of passwordless authentication [40, 63].

In the fast-moving field of IT security, staying up to date can be challenging – especially for non-experts who must adapt their online behaviors to an ever-evolving landscape. Changes such as automated updates or new authentication methods as passwordless authentication illustrate how security mechanisms have evolved, potentially influencing both user practices and expert recommendations. To examine whether and how these behaviors and pieces of advice have shifted, we replicated the studies by Ion et al. [46] and its replication by Busse et al. [17], providing an up-to-date perspective on IT security habits and guidance.

3 Method

We based our replication on materials published in prior work, as well as additional analysis materials, such as codebooks from Busse et al. [17] and the numbers from Figure 1 in the original study by Ion et al. [46], which the authors had subsequently shared with Busse et al. In the following, we describe the data collection and analysis process, highlighting differences to the published studies. We provide study materials, code books, analysis scripts and anonymized data on OSF to support further replication efforts. [1]

3.1 Positionality

Eight of the authors are part of a group doing research on usable security at a European university, including a professor and research group leader, and multiple PhD students and undergraduate students of computer science or IT security. Two other PhD student authors belong to different research groups, one at a North-American university in a group studying security and privacy, and one at a different European university, studying psychology of security and privacy. The group of authors collectively had experience with collecting and analyzing interview and survey data and conducting replication studies in UPS before this study. All of the authors have experienced situations in which they took on an expert role to give security advice to non-expert friends or family members. As such, their personal opinion on advice may have influenced

[1] https://osf.io/j43ws/?view_only=d2ab51af654e4cf78f35149f5a578fb7

the coding process of security practices and advice.

3.2 Ethics

The studies were approved by the Institutional Review Board (IRB) of our institution. To ensure compliant handling of participants' data and adherence to the General Data Protection Regulation (GDPR), we developed and followed a data management plan. All personal identifiers were either removed or pseudonymized to protect participant privacy. Data were stored on secure institutional servers with controlled access, deleted when no longer necessary, such as removing audio recordings after submission, and no personally identifiable information was shared beyond the research team.

3.3 Study Design

3.3.1 Expert Interviews

We conducted expert interviews at three security and privacy related events in Europe. The interviews were designed to take less than five minutes. All interviews began with the same question, to ensure consistency and comparability across all participants: "What are the top 3 pieces of advice you would give to a non-tech-savvy user to stay safe online?"

The follow-up questions varied between the interviews depending on the event. The first event took place before our replication of the survey study. The participants (N=18) were shown results from the non-expert surveys from the prior studies [17, 46], compiled into a single figure. Then, we asked for an open-ended description of their first impression, and if applicable followed up with questions on their opinion of the current status quo and possible future developments.

We used the results of these interviews to refine our survey. After collecting the results from the survey, we conducted more interviews at two further events with N=11and N=6 participants, respectively. This time, we showed participants the updated results, and again asked for their first impression. Follow-up questions asked for their opinion on the presented advice and for a perspective on possible changes in the future.

Following data minimization principles, we did not collect demographic information beyond participants' years of experience in IT security-related fields, as this was our main criterion for sample selection, and the source of their experience in IT security. We ensured that all participants had at least one year of experience.

3.3.2 Surveys

All participants gave informed consent before starting the survey. Non-expert participants then reported their top three security practices to stay safe online, and expert participants additionally submitted the top three pieces of advice they would give to a non-tech-savvy user. The second block of questions contained multiple-choice questions on participants' own security-related practices, and in the third block, participants rated security advice intended for a non-tech-savvy audience. Non-experts rated advice separately based on effectiveness and realism to implement the advice, while experts gave one overall rating of advice quality, incorporating effectiveness and realism. Finally, all participants answered demographic questions.

Based on the initial round of expert interviews and prior work on new authentication practices [40, 63, 97], we expanded the second and third survey block. We primarily added questions about new authentication methods such as hardware tokens, authenticator apps, passkeys and biometrics but also a question on sharing personal information, and checking permissions in mobile applications. In the third block, we added a question about using large language models (LLMs) such as ChatGPT to get security advice. Based on more recent best-practice, we adjusted the question on gender to include a non-binary option and the option to self-describe [79].

Finally, we added measures for bot detection. This was a necessary addition to our surveys, given the rising popularity of LLMs like ChatGPT [2], which can also be used by crowd workers to answer survey questions [20, 44, 47] and evade existing methods to detect them, such as CAPTCHAs [65]. We attempted to detect such responses by including several experimental methods for those participant groups which were paid for their participation. These methods included questions hidden through either font color or JavaScript, similar to measures that had been tested in prior work [45, 80].

Non-expert participants received direct compensation for their participation, and thus might be more incentivized to use tools like LLMs to efficiently answer surveys. To counter this, we used additional optional knowledge questions—which would be hard to answer for a human—as a bot detection measure. We explicitly stated that participants should not use other resources for these questions, and that they did not have to answer them.

For the expert survey, we reverted to using the compound question to ask experts for quality judgment of advice, as in the original study [46], since the additional questions about security behavior made the survey more lengthy overall.

3.4 Recruitment

We followed Busse et al. [17] in including security experts with at least one year of experience in an security related field.

For the interviews, we recruited security experts from three venues in Europe. The first was a UPS community event, which we choose not to name in this paper to prevent potential deanonymization of our participants. The other two venues were large enough that this concern does not arise, namely the BlackHat Europe, and the 38th Chaos Communication Congress (38C3), which are both hacker conferences.

[2] https://chatgpt.com/

Participants were not offered monetary compensation, but they received a stuffed toy as an incentive to take part.

For the non-expert surveys, we intended to recruit 300 non-expert participants on MTurk to replicate the prior studies as closely as possible, and additionally recruit a larger representative US-sample on Prolific to compare to the MTurk sample, given reports of declining data quality on MTurk [48, 82, 83]. However, due to the lack of data quality in our MTurk pilot sample (N=10), we elected to stop recruitment on MTurk and focus on Prolific instead. Non-expert participants were compensated with US-$3.

Similar to Busse et al. [17], we used social media to recruit for the expert survey. We published a call for participation on the most senior researcher's LinkedIn. We attempted to use different subreddits for recruitment, but mods of subreddits like *r/computerscience* and *r/sysadmin* had implemented rules against posting surveys or posts not strictly fitting the topic of the subreddit, so we were only able to post on *r/takemysurvey*. In addition, flyers were distributed at the interview venues. We also distributed our call for participation among contacts with a background in IT security and asked them to forward it further. We attempted to recruit additional security experts for a more diverse sample on Freelancer.com, which has been recommended for recruiting software developers in prior work [48]. To ensure data integrity, we added screening questions on basic security knowledge to the beginning of the survey. However, few developers passed these questions, leading to only four additional participants. Participants on Freelancer.com were compensated with 15€. All other expert participants were eligible to participate in a raffle for three times 100€.

3.5 Participants

Of the 35 interview participants, 18 were interviewed at the UPS community event (16 university, 1 industry, 1 government background), 11 at BlackHat Europe (9 industry, 1 university, 1 government background) and 6 at the 38C3 (3 university, 3 industry background).

A comparison of demographic data of the survey participants across the three iterations of the study is in Table 2, while more detailed participant information for the current iteration is in the appendix in Table 3. Following the prior studies, we excluded participants who made more than one mistake on the three attention checks in the non-expert survey (N=169), and two attention checks in the expert survey (N=2).

The non-expert survey originally had 997 participants after filtering based on attention checks. Six further participants were filtered out because they submitted the exact same answers to open questions and one participant did not complete the full survey, resulting in a sample of N=990.

Our 75 expert survey participants came from the following recruiting channels: 4 through flyers distributed at our expert interview venues, 18 through LinkedIn, 1 through Reddit, 48

	Ion et al. [46]	Busse et al. [17]	This paper
Expert Survey			
N	231	75	75
Gender			
Women	4 %	9 %	13 %
Men	~ %	79 %	84 %
Other	~ %	3 %	0 %
No answer	~ %	9 %	3 %
Age			
18 - 24	~ %	4 %	23 %
25 - 34	30 %	40 %	55 %
35 - 44	32 %	35 %	16 %
45 - 54	18 %	12 %	4 %
55 - 64	~ %	3 %	1 %
65 or older	~ %	0 %	1 %
No answer	~ %	7 %	0 %
Education			
Bachelor's or higher	73 %	75 %	85 %
Location			
US	47 %	27 %	5 %
Other	53 %	73 %	95 %
Work place			
Industry	69 %	51 %	37 %
University	15 %	21 %	43 %
Corporate research lab	7 %	9 %	5 %
Government	11 %	1 %	8 %
Self-employed	13 %	3 %	5 %
Other	~ %	~ %	1 %
Non-Expert Survey			
N	294	288	990
Gender			
Women	40 %	48 %	50 %
Men	~ %	52 %	48 %
Other	~ %	0 %	1 %
No answer	~ %	0 %	0 %
Age			
18 - 24	19 %	9 %	11 %
25 - 34	50 %	45 %	17 %
35 - 44	19 %	25 %	17 %
45 - 54	~ %	14 %	16 %
55 - 64	~ %	6 %	26 %
65 or older	~ %	2 %	12 %
No answer	~ %	0 %	0 %
Education			
Bachelor's or higher	47 %	52 %	57 %

Table 2: Demographic comparison for expert and non-expert demographics of Ion et al. [46], Busse et al. [17], and this paper. ~ indicates unknown percentages.

through personal contacts and 4 through Freelancer.com.

3.6 Limitations

Even though Busse et al's replication had shown that expert ratings for realism and effectiveness differ from responses to the compound question encompassing both [17], we reverted to Ion et al's original question phrasing [46]. Since we wanted to learn about experts' perceptions of new security practices, we had to balance effort and time needed for experts and the distinction between these two aspects of advice quality. The focus in this study was on new developments in security advice and as such we chose to accept the limitation of the compound question.

We attempted to replicate the non-expert survey recruitment on MTurk, following prior work. However, in our pilot sample of 10 participants, responses to open-ended questions appeared suspicious and contained patterns that suggested use of LLMs or bots, leading us to discontinue MTurk as a recruitment channel.

Like the prior replication [17] our expert sample is smaller than in the original study [46], and recruited from a different population, with a largely European expert sample with more university background. This might bear the risk of biasing the sample due to cultural differences, since Europe is widely believed to be generally more privacy conscious than the U.S. However, a recent study compared general population samples of both Germany and the U.S. and found that at least when it comes to government surveillance, there were only small differences in their privacy attitudes [38]. Using snowball sampling in the expert recruitment process means the responses may not be generalizable.

We based our coding of security practices and advice on the prior publications [17, 46], drawing from figures, tables, text, and additional coding material we received from the authors of the replication study. Nevertheless, qualitative coding is influenced by researchers' background and experiences [33, 73], and as such our application of the codes may have differed. In some instances, the prior iterations of the study used different names for codes. To increase readability, we unified these names when presenting the results.

3.7 Data Analysis

Given the wide range of advice evaluated and no clear hypotheses on expert vs. non-expert differences for specific advice, we chose descriptive statistics over null hypothesis testing. Nevertheless, to enable direct comparison to prior replications, we report on Fisher's exact tests comparing expert and non-expert behavior, and Wilcoxon signed rank test comparing non-experts' realism and effectiveness ratings, in the Tables 5 and 6 in the appendix. We report the effect sizes Cramer's V for the Fisher's Exact tests and rankbiserial r for the Wilcoxon signed rank tests, to enable comparisons even when sample sizes are different.

For qualitative analysis, we coded top-three responses (things-you-do for all, advice for experts) based on prior work [17, 46] and additional materials we received from Busse et al. After initial expert interviews, we expanded the codebook to include new advice, informing additional survey questions. Different sets of coders analyzed each of three types of data: Freetext answers from expert and non-expert surveys, and expert interviews. The main author took part in coding all types of data and ensured that the advice codebook was used consistently across data sets. For the survey answers, a subset of responses was coded independently by multiple coders, interrater reliability (IRR) was calculated and any differences in coding were discussed and resolved. This process was iteratively repeated until a sufficient IRR of at least $\kappa = 0.8$ had been achieved for the independent coding. The remaining data was split among coders. During this phase, coders marked responses for discussion if they were uncertain, and these were resolved together with all coders involved with the analysis of this type of data.

4 Results

We describe the findings from our re-replication, first discussing development of security behavior across the three iterations of the study, from 2014 to now. Next, we describe differences between experts and non-experts based on the surveys, focusing on newly included behavior and advice. Finally, based on the expert interviews, we describe an outlook into the future of security advice. We translated expert quotes to English where applicable. Quoted participants are represented by an identifier reflecting the interview venue: **CE** for the UPS community event, **BH** for BlackHat and **C3** for the 38C3.

4.1 Development of Security Behavior over Time (RQ1)

In our surveys, we asked both experts and non-experts to indicate the top 3 security enhancing behaviors that they carry out online. Figure 1 shows the development of experts' and non-experts' top 3 security behaviors across the three iterations of the survey. We only list behaviors which at least 10% of one sample (experts or non-experts in 2014, 2018 or 2024) named in their top 3, since there was a long tail of behaviors in the top 3 of barely over 5%.

We made some changes in naming the categories of the top 3 security practices, as compared to the works we build upon: Some respondents specified that they used multi-factor authentication, not only two-factor authentication, so we adjusted the code to include this: *Use 2FA/MFA*. Given the short responses, it is not clear if participants really use more than two factors or if they are using words they have encountered

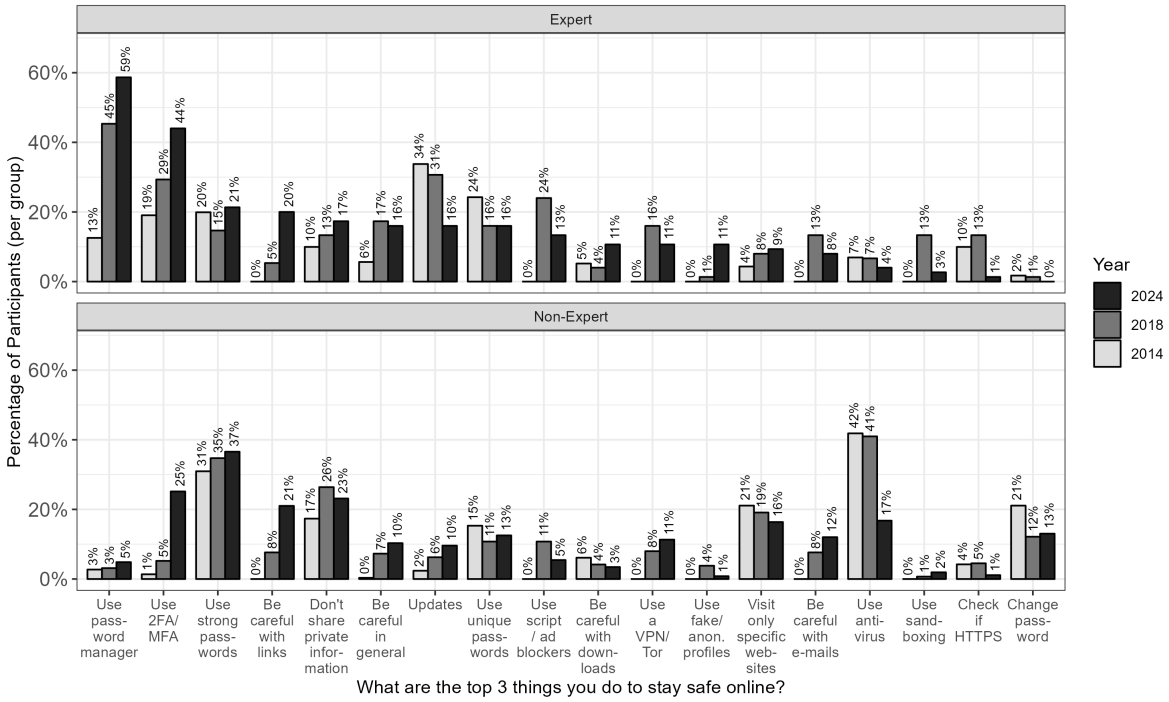


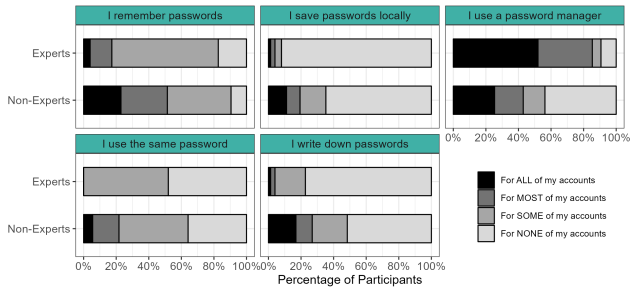
Figure 1: Behaviors from the top 3, where at least one sample had over 10%

during their practice. Similarly, many respondents did not clarify whether whether they meant updates on a (operating) system level, or of applications, so we summarized both as a single practice: *Updates*. We unified *Be careful/suspicious* to *Be careful in general* and use the description *Be careful with downloads* from [17] rather than *Use verified software* from the original study. Participants used other subjective criteria to decide which websites they would visit, including not only *known* and *trusted*, but also *professional* or *secure* websites, and avoided *risky* ones. To summarize these, we call the practice *Visit specific websites* and compare it to the practices *Visit known/trusted websites* from prior iterations. We shorten *Be careful with emails/attachments* to *Be careful with emails*.

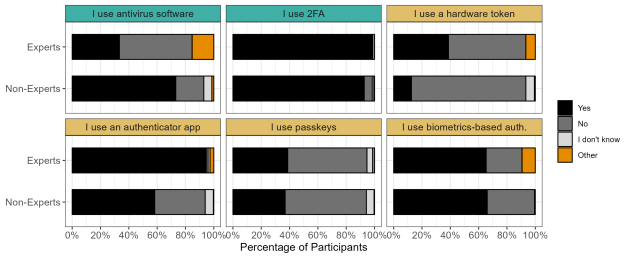
In Figure 1, some trends are immediately apparent, most notably the decrease in use of anti-virus software, and increase in use of 2FA. Also, non-experts increasingly report being careful with links. In the expert sample, the reported use of password managers and 2FA has increased, and the focus on using unique passwords and updating as security relevant behaviors has decreased. This is not to say that these measures are now deemed less important - they may also be seen as more self-evident or automated. Both the increased reported use of 2FA and the decrease in antivirus software was stronger in the non-expert sample compared to the expert sample. This leads to an overall more similar reported behavior between the two groups in the current sample, with a difference of 19 percent points in the use of 2FA as compared to the 24 percent points five years ago, and correspondingly 13 percent points compared to 34 in the use of antivirus.

We briefly compare this to current self-reported behavior (Figure 2b) and rating of these pieces of advice. 2FA was reported to be used for at least one account by both groups (E=99%, NE=93%). Non-experts rated the advice higher than experts. 44% of experts rated the advice as *very good*, while 74% of non-experts rated it as *very effective* and 66% considered themselves *very likely* to follow it. While using antivirus was among the top 3 self-reported security behaviors among non-experts in both 2014 and 2018, it now ranks fifth overall, with 17% of non-experts naming this in their top 3. Despite it being less popular than in previous years, non-experts still tend to be more favorable towards antivirus software than experts. While only 23% of experts rated using antivirus as *very good*, 50% of non-experts rated it *very effective* and 63% considered themselves *very likely* to follow the advice. Similarly, over twice as many non-experts reported using antivirus as experts (E=33%, NE=73%).

We identified seven new behaviors, which have not been discussed in prior work, but were reported by over 5% of either the expert or the non-expert sample: In our expert sample 11% considered *using fake/ anonymous profiles* online to stay anonymous among their top 3 security practices, while 1% of non-experts did this. Expert behaviors that were not included in Figure 1 were using *security settings*, e.g. in the browser or applications, *use encryption*, *use token-based authentication*, *self-host* and *limit internet use* (all >5%). For non-experts,



(a) Comparison of password behavior



(b) Comparison of antivirus and authentication behavior

Figure 2: Self-reported expert and non-expert security behavior, related to authentication practices. We use background color to mark whether behavior was newly introduced: yellow for new, blue for old.

one new security practice that had not been recorded in prior iterations was *use secure connections* (5%). This was often related to refraining from using public wifi, or at least being careful when using it.

4.2 Differences between Experts and Non-Experts (RQ2)

In the surveys, security advice was rated on a scale of 1 to 5. Experts were asked how good (combining effectiveness and realism) they would find the advice for a non-tech-savvy user. Non-experts were asked to indicate how likely it was that they would follow that same advice, and how effective they felt it would be. We asked experts and non-experts to indicate whether or not they engaged in various security practices. We first discuss differences in advice ratings between experts and non-experts. A visualization of the advice ratings discussed in the text is in Figure 3. We report the percentage of participants who rated advice as *good* (i.e., experts who rated it as *very good* or *good*), *effective* (i.e., non-experts who rated it as *very effective* or *effective*), and *realistic* (i.e., non-experts who stated that they are *very likely* or *likely* to follow the advice). When comparing over the whole range of the scale, we report means. An overview over the rating differences between experts and non-experts and breakdown of responses into specific rating levels is in Table 4 in the Appendix.

We also compare expert and non-expert behavior. Responses to the behavioral questions are in Figures 2, 4 and 5. Across all categories of practices, the response option *Other* was quite common for experts. They had more nuanced explanations on their behavior, which we will reference where appropriate in the following.

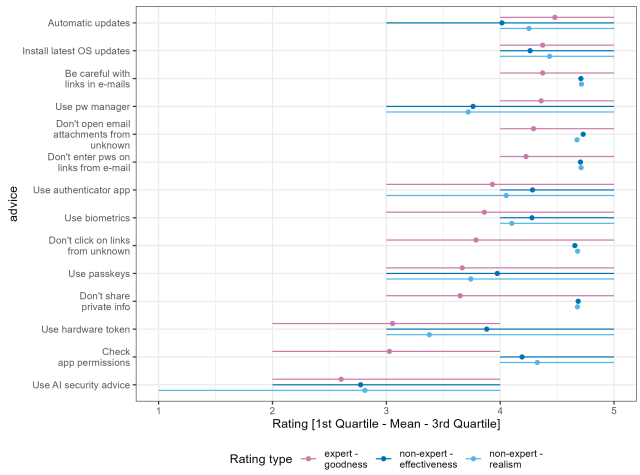


Figure 3: Means (dots) and Interquartile Range (lines) of experts' and non-experts' advice ratings for advice discussed in the text, sorted by expert ratings.

4.2.1 Rating of New Advice

We added new advice to be rated based both on the expert interviews at the UPS community event and new practices in [17] which had not previously been included in the advice to be rated. Expert and non-expert participants did not appear to be enthusiastic regarding using AI (such as ChatGPT) for security advice (Figure 3). Non-experts were more positive than experts about limiting sharing of unnecessary personal information (Figure 3), and also self-reporting that they did this (Figure 5). Furthermore, 65% of experts reported not sharing information, while 89% of non-experts did so, revealing diverging attitudes and practices between the two groups. However, this may be due to differences in what experts and non-experts determine to be "unnecessary" personal information or knowledge about what information is gathered and what can be inferred from it. This could potentially lead to higher rates of experts indicating they share such information. Two experts commented, one that they were not sure "what qualifies as unnecessary" and one that they restricted which people could access private information that they share. Meanwhile, though both groups check app permissions (often or sometimes) at similar rates (E=71%, NE=74%) (Figure 5), experts were much less positive about checking app permissions as security advice than NE (Figure 3).

Authentication Methods We also asked a set of new questions regarding authentication methods, and found that non-

experts tended to be less familiar with advice regarding authentication methods, in particular hardware tokens and passkeys. Figure 2b compares non-expert and expert behavior related to 2FA and use of different second factors and Figure 3 contains corresponding advice ratings.

Both experts (83%) and non-experts (87%) viewed securing accounts with 2FA positively, with high reported adoption (E= 99%, NE= 93%). Regarding the choice of the second factor, experts showed a clear preference for authentication apps, with 69% rating them as *good* advice, compared to only 34% for hardware tokens. This preference was reflected in their reported usage: 95% of experts reported using an authentication app for at least one account, whereas only 39% reported using hardware tokens. A similar pattern was observed among non-experts, though their overall reported adoption rate was lower: 58% reported using an authentication app, while only 13% used a hardware token. Interestingly, we found that non-experts struggled to understand hardware tokens, with 16% responding *I don't know* when asked about their effectiveness. However, non-experts appeared more familiar with passkeys than with hardware tokens. Rates of reported use of passkeys between experts and non-experts were similar (E and NE = 39%), though interestingly, as were rates of answering *I don't know* when asked if they used passkeys (E=4%, NE=5%). Despite the higher reported use rates of passkeys compared to hardware tokens for non-experts, this group did not tend to rate it as *effective* more frequently (64%) though there was an increase in the proportion of non-experts saying they would *likely* follow this advice (58%). Interestingly, experts appeared to be more uncertain about passkeys than non-experts, with more experts (24%) responding with *I don't know* regarding the goodness of advice on using passkeys than non-experts (11% *I don't know* for effectiveness, 6% *I don't know* for realism).

64% of experts thought using biometrics was *good* advice, while 80% of non-experts considered it *effective* and 73% said they were *likely* to follow this advice. Rates of use were similar between the two groups (E=65%, NE=66%). Experts responding *other* used biometrics to unlock their password manager or passkeys and were unsure whether this counted in the sense of our question. Two also explicitly mentioned using biometrics on their phones.

These results suggest that non-experts are less familiar with some of these authentication methods. It may also be that experts may feel that while the advice is effective (especially as a large percentage of these experts reported using these methods), it may not be realistic, which we were unable to distinguish given the compound format for our questions for experts. However, we did not see large discrepancies between the realism and effectiveness ratings from non-experts for this category of questions.

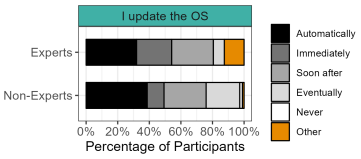


Figure 4: Comparison of expert and non-expert update behavior. This question was included in prior iterations of the study.

4.2.2 Remaining Usability Discrepancy

Password managers were consistently favored by experts, in terms of use and advice, but much less so by non-experts (Figure 2a and Figure 3). Experts were more positive towards advice on using password managers (56% *very good*) than non-experts (35% *very effective*, 41% *very likely*). It also is the most frequent of the top 3 actions that experts take to manage their own security (59%), though listed rarely by non-experts (5%) (Figure 1). The behavior of experts matches their evaluations of the advice, with 52% of experts reporting they use password managers for *all* of their accounts, while 25% of non-experts use it for *all* accounts. On the other hand, 44% of non-experts did not use password managers, as opposed to only 9% of experts. It was the only piece of advice rated as good (M=4.33) by experts but not realistic by non-experts (M=3.72). In fact, its realism score was the sixth lowest of all the pieces of advice according to non-experts, but the fourth best piece of advice from the mean rankings by experts. At the same time, expert interviewees acknowledged usability problems with password managers, e.g. P-BH-4 explained "if [...] you set it up correctly. It actually works really really well. But [...] sometimes I'm like why the hell is this not working on my mobile phone?" Thus, we see a mismatch between what experts believe non-experts should do in terms of security, and what non-experts believe.

4.2.3 Best Rated Advice

We find discrepancies between what experts and non-experts think is good advice, and how much the advice is self-reported to be followed in practice by both experts and non-experts. A visualization of the rating differences is in Figure 3.

Experts' Best Rated Advice Experts ranked advice relating to software and OS updates highly, with two pieces of advice related to updates in the top 3 in terms of mean ratings of advice goodness. Automatic software updates was ranked the "best" piece of advice by experts in terms of mean goodness rating (4.48). Experts were somewhat more favorable towards this advice than non-experts. Despite being more positive than non-experts towards automatic software updates, fewer experts said that OS updates were installed automatically than non-experts (E=32%, NE=39%) (Figure 4). Nonetheless, experts appear to be a little more timely at installing OS updates

in general than non-experts. This aligns with the advice to install the latest OS updates, which was tied for second rank in terms of mean goodness (4.37) according to experts. Regarding self-reported behavior, 81% of experts reported that they either updated their OS *automatically*, *immediately*, or *soon after* (with a further 13% responding with *other*), while a slightly lower proportion of non-experts (76%) install the latest OS updates. This may be because while experts value keeping their OS updated, they prefer to have more control over their own OS, and thus think of automatic updates as good for non-expert users rather than for themselves. Among the 13% of experts providing freetext answers to the question on their OS update behavior, some experts discussed reasons for their update frequency, e.g. delaying updates to monitor stability or updating on a rolling release base.

In some cases, experts and non-experts had similar evaluations about what is good security advice, as both groups ranked the advice to be suspicious of links in emails and messages highly. It had the second highest (though tied) mean rating for goodness (4.37) according to experts as well as the highest for realism (4.71) and second highest for effectiveness (4.71) for non-experts.

Non-experts' Best Rated Advice The advice to not enter passwords on sites linked in emails is rated the second highest in realism (M=4.71) and third highest in effectiveness (M=4.71) for non-experts. In fact, 77% of non-experts say they follow this advice and never enter their passwords on sites linked from emails (Figure 5). Experts were less positive about this advice. While 77% of experts rated the advice as *good*, a smaller percent of experts (60%) say they follow this advice. 20% of experts respond with *other*, suggesting more nuance to expert practices than a clear-cut yes or no. Not opening email attachments from unknown senders ranked the highest in mean effectiveness for non-experts (4.73). It is followed by 56% of non-experts. Despite 80% of experts rating it as good advice, only 11% of experts (with a further 6% of experts answering *other*) actually follow it (Figure 5). Not clicking on links sent by unknown senders was rated by non-experts as the third most realistic advice (M=4.68). Experts were not as positive (61% = *good*) towards this advice compared to non-experts (rated *effective* and *realistic* by 92% each of non-experts), nor did they report to follow it in practice (65%) as much as non-experts (82%) (Figure 5).

From the trends in the most highly rated advice from experts and non-experts, we see diverging opinions of what is best for one's security. In general, both groups believe that being aware of potential risks (e.g., be suspicious of links in emails and messages) is effective and realistic. However, the groups differ in what specific actions to take (or not take). The results suggest that experts think *positive* security advice (e.g., installing updates) to take an action that increases security is good, while non-experts think *negative* security advice (e.g., not opening emails, not entering passwords on unknown sites) to not take an action that potentially decreases security are most effective or realistic.

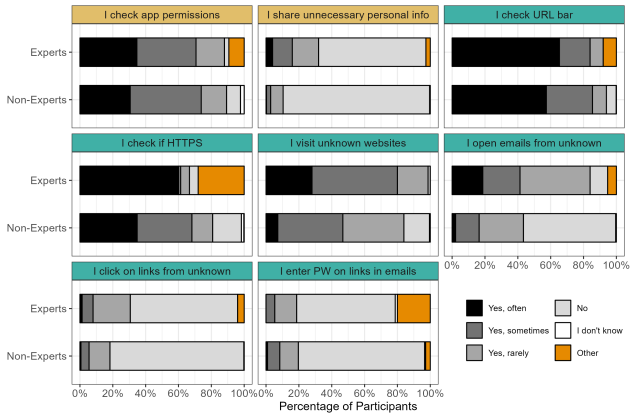


Figure 5: Self-reported expert and non-expert security behavior, related to mindfulness. We use background color to mark whether behavior was newly introduced: yellow for new, blue for old.

4.3 Experts' View on the Current State of End User Security Advice Practices (RQ3)

4.3.1 Security By Default

Many interviewed experts felt that IT security should not be the responsibility of end users. They discussed examples of how responsibility for security can be kept away from users, some of which are already implemented and others they would like to see in the future. This included the automation of software updates, use of anti-virus software, and link security.

A common view was that keeping software up to date is crucial for security. Many experts supported automatic updates, as they reduce the users' responsibility. However, both expert discussion and usable security research emphasize the need for clear communication and some user control to avoid frustration [93]. As one interviewee put it, "It's simply updated whether you like it or not. [...] so I wouldn't have to put so much effort into it, because there are simply updates" (P-CE-11). A similar view was expressed in relation to anti-virus protection. Some experts pointed out that solutions such as Windows Defender have improved significantly, making third-party anti-virus software unnecessary for most users. As one expert put it, "Use antivirus, oh yes, and thankfully this is a default in Windows" (P-CE-18). Link security was another topic that came up frequently in the interviews. Several experts noted that links are meant to be interacted with, so it is important that browsers and email clients provide better protection. As P-CE-18 explained, "Because links are made to be clicked overall, so it's about the infrastructure to help with that, [...] the browser or the email client". This is a topic where experts still saw the need for improvement. Overall, many of the experts interviewed shared the view that security

should be built into systems by default, reducing the need for user awareness and manual intervention. However, there was also an emphasis on maintaining transparency and giving users a degree of control to avoid frustration.

4.3.2 Risks and Potential of AI

When speaking of AI, our expert interviewees largely discussed it in terms of risks. These risks referred to privacy, with user data being used to train models, when users interact with AI and AI-powered services. One interviewee also worried about AI-driven possibilities for deepfakes. These concerns lead the experts to advise strongly that users avoid sharing personal information, specifically with AI, but also more generally online, given that training data can also stem from social media and the broader internet [9]. These expert views are in line with non-experts' views from a recent multi-country survey study [34]. However, P-BH-4 also sees chances in the use of AI: "Threat actors are using AI, but defenders need to actually get better at using AI."

4.3.3 Focused and Specific Advice

Some interviewees pointed out gaps in the advice presented from prior and current results based on non-expert data, for example in the way that advice seemed overly geared towards computers: "This is a time, where in many countries almost nobody has a computer anymore but they are working with tablets, they are working with smartphones. And adding to that all the verbal devices [...]" (P-CE-6) The interviewees saw the need for device and situation-specific advice, e.g., advice focused on mobile or internet of things devices. This rise in different types of devices may provide chances to aid secure behavior online, for example, using encrypted messaging apps was perceived as a low-barrier method to communicate securely on mobile devices. But interviewees also pointed out risks, e.g., an increased risk of shoulder surfing when using mobile devices in public spaces, or lack of transparency for data transmission processes on internet of things devices.

Interviewees also mentioned the need for advice being focused on specific user groups. The interviewees saw some technological solutions improving security, like password managers or 2FA, as complex for non-experts, but more accessible for experts: "Use of firewall is irrelevant for most people. They have a router in the back, then they don't need anything else" (P-C3-4).

5 Discussion

5.1 Changing Patterns of Security Practices

Over the ten years since the original "No one can hack my mind" study was published, we observe development in non-expert and expert security practices. In this series of studies, new advice appeared in both replications, e.g., using script and ad blockers or VPNs in 2018 and using security settings or fake profiles in 2024. Non-experts have adopted and value security practices that were previously mainly used by experts. In 2024, the most extreme change was in using 2FA. While this was among the pieces of advice that were rated as effective by experts in 2018, now it is the second most frequently named behavior that is seen as important by non-experts, with 97% of non-experts reporting using it for at least one account. Adoption of security technology, like 2FA, can be driven by usability improvements [97] but mandating the technology use may also be an influencing factor [1, 23]. Mandates are however a double-edged sword and may lead to security fatigue [1, 23, 62], and difficulties in deviating from in-grained practices, should advice change over the course of time, as was the case for recommendations on regularly changing passwords [39].

On the other hand, some security practices decreased in importance, such as the use of anti-virus software in the perception of non-experts, and updates in the perception of experts. In our expert interviews, automation was seen as a reason that these practices may have become less visible. Since they are often activated by default, their users may not be aware of them and thus their importance is sidelined. This aligns with a long-standing debate in the USP community about the trade-off between automation and user control [21, 26, 55, 88]. Automating security responsibility away from users also carries the risk of lack of transparency and feelings of loss of control and trust, leading to users circumventing automated security [29, 53, 88, 93]. Also, automation requires predefined rules for decisions: Their accuracy can be limited within flexible social situations and changing contexts, resulting in failures of the automation and, thus, involvement of users [26]. In turn, security design should balance automation with clear communication, minimizing responsibility for non-technical users by implementing secure defaults [62] while retaining control for those willing to exert it and keeping users in the loop [21, 26]. We propose to continue to monitor the state of advice perception among non-experts and experts to identify challenges and further fields of action.

5.2 Divergences Between Experts and Non-Experts

While for some advice, expert and non-expert behavior became more similar, there were still many differences between the two groups. One of the most striking was the attitude towards using password managers, which was the single most frequently reported behavior within experts' top 3, and received the fourth best rating. However, this advice was rated as unrealistic by non-experts with the sixth lowest mean rating. This can be considered a remaining challenge for the UPS community. On the other hand, some advice like not clicking on links from unknown sources was perceived as

realistic by non-experts, but not good advice by experts. Additionally, even experts are not consistent in their perception of advice. Two of the top 3 pieces of advice about updates received high ratings from experts, but were not perceived as realistic by non-experts and experts did not really follow the advice themselves. A prior study from 2018 similarly found that self-declared experts reported fewer security behaviors [18]. However, this likely stems from experts' more nuanced view on advice, which they explained in free text answers. They have the confidence or knowledge to deviate from advice when they see it as not necessary, but nevertheless experts themselves also diverge in their behavior and recommendations, as shown in related work [69].

5.3 Advice Should Be Context-Specific

While the primary question in this study asks about "staying safe online", the meaning of online has changed in the past ten years. The internet is now more than ever used through and by multiple devices and in different contexts, from internet cafés in Kenya [57] to smart home environments encompassing interactions between multiple devices and multiple people [59, 77], and connected cars [14]. Similar to these fragmented online use patterns, the status quo of security advice is also confusing for users, with a wide variety of advice available and experts not able to agree on prioritization of advice, leaving non-experts confused and without clear actionable advice [69]. Perhaps the way forward is not to try and find one set of the universal best pieces of advice, but rather identify the advice most salient for individual users or user groups. UPS research has already embarked on this endeavor by investigating advice for older adults [56], activists [15], LGBTQ+ folks [37] or software developers [2], and how, where and why it is circulated and implemented or dismissed. For other user groups, such as users with visual impairments, their difficulties in achieving security and privacy online have been documented [42, 58], but there has been no specific investigation of advice directed at these users.

Another way to focus advice is by providing advice for specific devices. Two examples mentioned in our expert interviews were mobile devices and internet of things devices. Research on mobile security focuses on specific security or privacy related functionality, such as permission management [4, 66, 81] or authentication [19, 41], but there is no evaluation of mobile-focused advice. Much of the available security advice for internet-of-things devices was not actionable, and not understandable for non-expert users [10, 86].

6 Reflections on Participant Recruitment

Recruiting security experts for empirical research remains a persistent methodological challenge. These individuals represent a hard-to-reach population, particularly for large-scale, survey-based studies [48]. In our study, we used various recruitment strategies to target experts. Personal networks proved to be the most reliable channel, similar to findings that software developer participants preferred personal contact for being recruited [74], though this approach is likely to introduce sample selection biases. Verifying participants' expertise poses an additional challenge, as non-experts are financially incentivised to pose as experts. This problem is compounded by the growing sophistication of LLMs, which lower the barrier to deception. This may explain the low success rate of our recruitment efforts on Freelancer.com. Though ChatGPT-resistant screening instruments exists for participants with programming expertise [75], there are varied forms of security expertise, making it hard to pinpoint and verify. As screening becomes a cat-and-mouse game, future studies may need to rely more heavily on settings where participants' identity and expertise can be verified through direct interaction.

7 Conclusion

Ten years after Ion et al.'s study comparing experts' and non-experts' security-related behavior and advice, we conducted a second replication study to monitor the development of security practices over time. In addition to the replication, we investigated how newer advice, such as more recent advances in passwordless authentication, are perceived by experts and non-experts. We found that experts and non-experts are beginning to adopt new security practices, but overall experts' and non-experts' security practices continue to differ. Ten years are a period that yields relevant change in security behavior both for experts and non-experts and we call for longitudinal studies monitoring such developments in UPS.

Acknowledgments

The authors would like to thank Carina Lübcke and Charlotte Starke for their contributions to recruiting expert survey participants, Charlotte Mädler for her help in interviewing, and Ben Weinshel for initial discussions surrounding MTurk recruitment. This work was partially funded by the Werner Siemens Foundation.

References

[1] Jacob Abbott and Sameer Patil. How Mandatory Second Factor Affects the Authentication User Experience. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, New York, USA, 2020. ACM.

[2] Yasemin Acar, Christian Stransky, Dominik Wermke, Charles Weir, Michelle L. Mazurek, and Sascha Fahl. Developers Need Support, Too: A Survey of Security Advice for Software Developers. In *Proceedings of the*

2017 IEEE Cybersecurity Development, pages 22–26, Cambridge, USA, 2017. IEEE.

[3] Anne Adams and Martina Angela Sasse. Users are not the enemy. *Communications of the ACM*, 42(12):40–46, 1999. Publisher: ACM ISBN: 0001-0782.

[4] Mamtaj Akter, Madiha Tabassum, Nazmus Sakib Miazi, Leena Alghamdi, Jess Kropczynski, Pamela J. Wisniewski, and Heather Lipford. Evaluating the Impact of Community Oversight for Managing Mobile Privacy and Security. In *Proceedings of the Nineteenth Symposium on Usable Privacy and Security*, pages 437–456, Anaheim, USA, 2023. USENIX Association.

[5] Ahmed AlEroud and George Karabatis. Bypassing Detection of URL-based Phishing Attacks Using Generative Adversarial Deep Neural Networks. In *Proceedings of the Sixth International Workshop on Security and Privacy Analytics*, pages 53–60, New York, USA, 2020. ACM.

[6] Zainab Alkhalil, Chaminda Hewage, Liqaa Nawaf, and Imtiaz Khan. Phishing Attacks: A Recent Comprehensive Study and a New Anatomy. *Frontiers in Computer Science*, 3, 2021.

[7] Sabrina Amft, Sandra Höltervennhoff, Nicolas Huaman, Yasemin Acar, and Sascha Fahl. "Would You Give the Same Priority to the Bank and a Game? I Do Not!" Exploring Credential Management Strategies and Obstacles during Password Manager Setup. In *Proceedings of the Nineteenth Symposium on Usable Privacy and Security*, page 21, Anaheim, USA, 2023. USENIX Association.

[8] Apple.com. Security. Built right in. https://web.archive.org/web/20250124212613/https://www.apple.com/macos/security/, November 2020.

[9] Stefan Baack. A Critical Analysis of the Largest Source for Generative AI Training Data: Common Crawl. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, pages 2199–2208, Rio de Janeiro, Brazil, 2024. ACM.

[10] David Barrera, Christopher Bellman, and Paul Van Oorschot. Security Best Practices: A Critical Analysis Using IoT as a Case Study. *ACM Trans. Priv. Secur.*, 26(2):13:1–13:30, 2023.

[11] Sruti Bhagavatula, Lujo Bauer, and Apu Kapadia. "Adulthood is trying each of the same six passwords that you use for everything": The Scarcity and Ambiguity of Security Advice on Social Media. *Proc. ACM Hum.-Comput. Interact.*, 6(CSCW2):264:1–264:27, 2022.

[12] Joseph Bonneau, Cormac Herley, Paul C. van Oorschot, and Frank Stajano. The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes. In *2012 IEEE Symposium on Security and Privacy*, pages 553–567, 2012. ISSN: 2375-1207.

[13] Joseph Bonneau, Cormac Herley, Paul C. van Oorschot, and Frank Stajano. Passwords and the evolution of imperfect authentication. *Communications of the ACM*, 58(7):78–87, 2015.

[14] Paul Bossauer, Thomas Neifer, Gunnar Stevens, and Christina Pakusch. Trust versus Privacy: Using Connected Car Data in Peer-to-Peer Carsharing. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, New York, USA, 2020. ACM.

[15] Maia J. Boyd, Jamar L. Sullivan Jr., Marshini Chetty, and Blase Ur. 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*, New York, USA, 2021. ACM.

[16] Kay Bryant and John Campbell. User Behaviours Associated with Password Security and Management. *Australasian Journal of Information Systems*, 14(1), 2006.

[17] Karoline Busse, Julia Schäfer, and Matthew Smith. Replication: No one can hack my mind revisiting a study on expert and non-expert security practices and advice. In *Fifteenth symposium on usable privacy and security*, Santa Clara, CA, 2019. USENIX Association.

[18] Ashley A. Cain, Morgan E. Edwards, and Jeremiah D. Still. An exploratory study of cyber hygiene behaviors and knowledge. *Journal of Information Security and Applications*, 42:36–45, 2018.

[19] Geumhwan Cho, Sungsu Kwag, Jun Ho Huh, Bedeuro Kim, Choong-Hoon Lee, and Hyoungshick Kim. Towards Usable and Secure Location-based Smartphone Authentication. In *Proceedings of the Seventeenth Symposium on Usable Privacy and Security*. USENIX Association, 2021.

[20] Evgenia Christoforou, Gianluca Demartini, and Jahna Otterbacher. Generative AI in Crowdwork for Web and Social Media Research: A Survey of Workers at Three Platforms. *Proceedings of the International AAAI Conference on Web and Social Media*, 18:2097–2103, 2024.

[21] Lorrie Faith Cranor. A Framework for Reasoning About the Human in the Loop. *Proceedings of the 1st Conference on Usability, Psychology, and Security*, 2008.

[22] Anupam Das, Joseph Bonneau, Matthew Caesar, Nikita Borisov, and XiaoFeng Wang. The Tangled Web of Password Reuse. In *Proceedings 2014 Network and Distributed System Security Symposium*, San Diego, CA, 2014. Internet Society.

[23] Sanchari Das, Andrew Kim, Shrirang Mare, Joshua Streiff, and L. Jean Camp. Security Mandates are Pervasive: An Inter-School Study on Analyzing User Authentication Behavior. In *2019 IEEE 5th International Conference on Collaboration and Internet Computing*, pages 306–313. IEEE, 2019.

[24] Rachna Dhamija, J. D. Tygar, and Marti Hearst. Why phishing works. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 581–590, Montréal Québec Canada, 2006. ACM.

[25] Thomas Duebendorfer and Stefan Frei. Why Silent Updates Boost Security.

[26] W. Keith Edwards, Erika Shehan Poole, and Jennifer Stoll. Security automation considered harmful? In *Proceedings of the 2007 Workshop on New Security Paradigms*, pages 33–42, New Hampshire, 2008. ACM.

[27] Serge Egelman, Lorrie Faith Cranor, and Jason Hong. You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1065–1074, Florence Italy, 2008. ACM.

[28] Michael Fagan and Mohammad Maifi Hasan Khan. Why do they do what they do?: A study of what motivates users to (not) follow computer security advice. In *Proceedings of the Twelfth symposium on usable privacy and security*, pages 59–75, Denver, CO, 2016. USENIX Association.

[29] Michael Fagan, Mohammad Maifi Hasan Khan, and Ross Buck. A study of users' experiences and beliefs about software update messages. *Computers in Human Behavior*, 51:504–519, 2015.

[30] Matthias Fassl, Michaela Neumayr, Oliver Schedler, and Katharina Krombholz. Transferring Update Behavior from Smartphones to Smart Consumer Devices. In Sokratis Katsikas, Costas Lambrinoudakis, Nora Cuppens, John Mylopoulos, Christos Kalloniatis, Weizhi Meng, Steven Furnell, Frank Pallas, Jörg Pohle, M. Angela Sasse, Habtamu Abie, Silvio Ranise, Luca Verderame, Enrico Cambiaso, Jorge Maestre Vidal, and Marco Antonio Sotelo Monge, editors, *Computer Security. ESORICS 2021 International Workshops*, pages 357–383, Cham, 2022. Springer International Publishing.

[31] Dinei Florencio and Cormac Herley. A large-scale study of web password habits. In *Proceedings of the 16th international conference on World Wide Web*, pages 657–666, Banff Alberta, Canada, 2007. ACM.

[32] Alain Forget, Sarah Pearman, Jeremy Thomas, Alessandro Acquisti, Nicolas Christin, Lorrie Faith Cranor, Serge Egelman, Marian Harbach, and Rahul Telang. Do or do not, there is no try: user engagement may not improve security outcomes. In *Proceedings of the Twelfth USENIX Conference on Usable Privacy and Security*, pages 97–111, USA, 2016. USENIX Association.

[33] Nollaig Frost, Sevasti Melissa Nolas, Belinda Brooks-Gordon, Cigdem Esin, Amanda Holt, Leila Mehdizadeh, and Pnina Shinebourne. Pluralism in qualitative research: the impact of different researchers and qualitative approaches on the analysis of qualitative data. *Qualitative Research*, 10(4):441–460, 2010.

[34] Patrick Gage Kelley, Celestina Cornejo, Lisa Hayes, Ellie Shuo Jin, Aaron Sedley, Kurt Thomas, Yongwei Yang, and Allison Woodruff. "There will be less privacy, of course": How and why people in 10 countries expect AI will affect privacy in the future. In *Proceedings of the Nineteenth Symposium on Usable Privacy and Security*, Anaheim, USA, 2023. USENIX Association.

[35] Faisal A. Garba, Rosemary M. Dima, A. Balarabe Isa, A. Abdulrazaq Bello, A. Sarki Aliyu, F. Umar Yarima, and S. Abbas Ibrahim. Re-Evaluating the Necessity of Third-Party Antivirus Software on Windows Operating System. *Journal of Cybersecurity and Information Management*, 10(1):18–33, 2022.

[36] Shirley Gaw and Edward W. Felten. Password management strategies for online accounts. In *Proceedings of the second symposium on Usable privacy and security*, pages 44–55, New York, USA, 2006. ACM.

[37] Christine Geeng, Mike Harris, Elissa Redmiles, and Franziska Roesner. "Like Lesbians Walking the Perimeter": Experiences of {U.S}. {LGBTQ+} Folks With Online Security, Safety, and Privacy Advice. In *Proceedings of the 31st USENIX Security Symposium*, pages 305–322, Boston, USA, 2022. USENIX Association.

[38] Lisa Geierhaas, Florin Martius, Arthi Arumugam, and Matthew Smith. "Not the Right Question?" A Study on Attitudes Toward Client-Side Scanning with Security and Privacy Researchers and a US Population Sample. In *2025 IEEE Symposium on Security and Privacy*. IEEE Computer Society, 2024.

[39] Eva Gerlitz, Maximilian Häring, Matthew Smith, and Christian Tiefenau. Evolution of Password Expiry in

Companies: Measuring the Adoption of Recommendations by the German Federal Office for Information Security. In *Nineteenth Symposium on Usable Privacy and Security*, pages 191–210, Anaheim, USA, 2023. USENIX Association.

[40] Sanam Ghorbani Lyastani, Michael Schilling, Michaela Neumayr, Michael Backes, and Sven Bugiel. Is FIDO2 the Kingslayer of User Authentication? A Comparative Usability Study of FIDO2 Passwordless Authentication. In *2020 IEEE Symposium on Security and Privacy*, pages 268–285, San Francisco, USA, 2020. IEEE.

[41] Marian Harbach, Emanuel von Zezschwitz, Andreas Fichtner, Alexander De Luca, and Matthew Smith. It's a Hard Lock Life: A Field Study of Smartphone (Un)Locking Behavior and Risk Perception. In *Proeedings of the 10th Symposium On Usable Privacy and Security*, pages 213–230, Menlo Park, USA, 2014. USENIX Association.

[42] Jordan Hayes, Smirity Kaushik, Charlotte Emily Price, and Yang Wang. Cooperative Privacy and Security: Learning from People with Visual Impairments and Their Allies. In *Proceedings of the Fifteenth Symposium on Usable Privacy and Security*. USENIX Association, 2019.

[43] Cormac Herley. So long, and no thanks for the externalities: the rational rejection of security advice by users. In *Proceedings of the 2009 workshop on New security paradigms workshop*, pages 133–144, New York, USA, 2009. ACM.

[44] Perttu Hämäläinen, Mikke Tavast, and Anton Kunnari. Evaluating Large Language Models in Generating Synthetic HCI Research Data: a Case Study. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, Hamburg, Germany, 2023. ACM.

[45] Jan Karem Höhne, Joshua Claassen, Saijal Shahania, and David Broneske. Bots in web survey interviews: A showcase. *International Journal of Market Research*, 67(1):3–12, 2025. Publisher: SAGE Publications.

[46] Iulia Ion, Rob Reeder, and Sunny Consolvo. "...No one can hack my mind": Comparing expert and non-expert security practices. In *Proceedings of the Eleventh symposium on usable privacy and security*, pages 327–346, Ottawa, 2015. USENIX Association.

[47] Bernard J. Jansen, Soon-gyo Jung, and Joni Salminen. Employing large language models in survey research. *Natural Language Processing Journal*, 4:100020, 2023.

[48] Harjot Kaur, Sabrina Klivan, Daniel Votipka, Yasemin Acar, and Sascha Fahl. Where to Recruit for Security Development Studies: Comparing Six Software Developer Samples. In *Proceedings of the 31st USENIX Security Symposium*, pages 4041–4058, Boston, USA, 2022. USENIX Association.

[49] Katharina Krombholz, Karoline Busse, Katharina Pfeffer, Matthew Smith, and Emanuel Von Zezschwitz. "If HTTPS Were Secure, I Wouldn't Need 2FA" - End User and Administrator Mental Models of HTTPS. In *2019 IEEE Symposium on Security and Privacy*, pages 246–263, San Francisco, USA, 2019. IEEE.

[50] Sanam Ghorbani Lyastani, Michael Schilling, Sascha Fahl, Sven Bugiel, and Michael Backes. Studying the Impact of Managers on Password Strength and Reuse, 2017. arXiv:1712.08940 [cs].

[51] Zane Ma, Joshua Reynolds, Joseph Dickinson, Kaishen Wang, Taylor Judd, Joseph D. Barnes, Joshua Mason, and Michael Bailey. The Impact of Secure Transport Protocols on Phishing Efficacy. In *12th USENIX Workshop on Cyber Security Experimentation and Test*, Santa Clara, USA, 2019. USENIX Association.

[52] Paolo Matarazzo and MokumaPM. Operating System Security. https://web.archive.org/web/20241207151426/https://learn.microsoft.com/en-us/windows/security/book/operating-system-security, November 2024.

[53] Arunesh Mathur and Marshini Chetty. Impact of User Characteristics on Attitudes Towards Automatic Mobile Application Updates. In *Proceedings of the Thirteenth Symposium on Usable Privacy and Security*, pages 175–193. USENIX Association, 2017.

[54] Arunesh Mathur, Nathan Malkin, Marian Harbach, Eyal Peer, and Serge Egelman. Quantifying Users' Beliefs about Software Updates. In *Proceedings 2018 Workshop on Usable Security*, 2018. arXiv:1805.04594 [cs].

[55] Raydel Montesino and Stefan Fenz. Information Security Automation: How Far Can We Go? In *2011 Sixth International Conference on Availability, Reliability and Security*, pages 280–285, Vienna, Austria, 2011. IEEE.

[56] Benjamin Morrison, Lynne Coventry, and Pam Briggs. How do Older Adults feel about engaging with Cyber-Security? *Human Behavior and Emerging Technologies*, 3(5):1033–1049, 2021.

[57] Collins W. Munyendo, Yasemin Acar, and Adam J. Aviv. "In Eighty Percent of the Cases, I Select the Password for Them": Security and Privacy Challenges, Advice, and Opportunities at Cybercafes in Kenya. In *2023 IEEE Symposium on Security and Privacy*, pages 570–587, San Francisco, USA, 2023. IEEE.

[58] Daniela Napoli, Khadija Baig, Sana Maqsood, and Sonia Chiasson. "I'm Literally Just Hoping This Will Work": Obstacles Blocking the Online Security and Privacy of Users with Visual Disabilities. In *Proceedings of the Seventeenth Symposium on Usable Privacy and Security*, pages 263–280. USENIX Association, 2021.

[59] Sangeun Oh, Hyuck Yoo, Dae R. Jeong, Duc Hoang Bui, and Insik Shin. Mobile Plus: Multi-device Mobile Platform for Cross-device Functionality Sharing. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, pages 332–344, New York, USA, 2017. ACM.

[60] Hrithik Padalia, Hitesh Patel, Amarjit Deshmukh, Mahadev Patil, Ajay Kumar, and Nripesh Kumar Nrip. A Study on Password Manager: Users' Perspective. In *2023 International Conference on Computational Intelligence for Information, Security and Communication Applications*, pages 72–75, 2023.

[61] Thomas Kobber Panum, Kaspar Hageman, René Rydhof Hansen, and Jens Myrup Pedersen. Towards Adversarial Phishing Detection. In *13th USENIX Workshop on Cyber Security Experimentation and Test*. USENIX Association, 2020.

[62] Simon Parkin, Kat Krol, Ingolf Becker, and M. Angela Sasse. Applying Cognitive Control Modes to Identify Security Fatigue Hotspots. In *Proceedings of the Twelfth Symposium on Usable Privacy and Security*, Security Fatigue Workshop'16. USENIX Association, 2016.

[63] Viral Parmar, Harshal A. Sanghvi, Riki H Patel, and Abhijit S. Pandya. A Comprehensive Study on Passwordless Authentication. In *2022 International Conference on Sustainable Computing and Data Communication Systems*, pages 1266–1275, 2022.

[64] Katharina Pfeffer, Alexandra Mai, Adrian Dabrowski, Matthias Gusenbauer, Philipp Schindler, Edgar Weippl, Michael Franz, and Katharina Krombholz. On the Usability of Authenticity Checks for Hardware Security Tokens. In *Proceedings of the 30th USENIX Security Symposium*, online, 2021. USENIX Association.

[65] Andreas Plesner, Tobias Vontobel, and Roger Wattenhofer. Breaking reCAPTCHAv2. In *2024 IEEE 48th Annual Computers, Software, and Applications Conference*, pages 1047–1056, 2024. ISSN: 2836-3795.

[66] Sarah Prange, Pascal Knierim, Gabriel Knoll, Felix Dietz, Alexander De Luca, and Florian Alt. "I do (not) need that Feature!" – Understanding Users' Awareness and Control of Privacy Permissions on Android Smartphones. In *Proceedings of the Twentieth Symposium on Usable Privacy and Security*, pages 453–472, Philadelphia, USA, 2024. USENIX Association.

[67] Elissa M. Redmiles, Sean Kross, and Michelle L. Mazurek. How I Learned to be Secure: a Census-Representative Survey of Security Advice Sources and Behavior. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 666–677, Vienna, Austria, 2016. ACM.

[68] Elissa M. Redmiles, Amelia R. Malone, and Michelle L. Mazurek. I Think They're Trying to Tell Me Something: Advice Sources and Selection for Digital Security. In *Proceedings of the 2016 IEEE Symposium on Security and Privacy (SP)*, pages 272–288, San Jose, USA, 2016. IEEE.

[69] Elissa M. Redmiles, Noel Warford, Amritha Jayanti, Aravind Koneru, Sean Kross, Miraida Morales, Rock Stevens, and Michelle L. Mazurek. A comprehensive quality evaluation of security and privacy advice on the web. In *Proceedings of the 29th USENIX security symposium*, pages 89–108. USENIX Association, 2020.

[70] Robert W. Reeder, Iulia Ion, and Sunny Consolvo. 152 Simple Steps to Stay Safe Online: Security Advice for Non-Tech-Savvy Users. *IEEE Security & Privacy*, 15(5):55–64, 2017. Conference Name: IEEE Security & Privacy.

[71] Anna Lena Rotthaler, Harshini Sri Ramulu, Lucy Simko, Sascha Fahl, and Yasemin Acar. "It's time. Time for digital security.": An End User Study on Actionable Security and Privacy Advice. In *IEEE Symposium on Security and Privacy*, page 100. IEEE Computer Society, 2024. ISSN: 2375-1207.

[72] Ozgur Koray Sahingoz, Ebubekir Buber, Onder Demir, and Banu Diri. Machine learning based phishing detection from URLs. *Expert Systems with Applications*, 117:345–357, 2019.

[73] Shruti Sannon and Andrea Forte. Privacy Research with Marginalized Groups: What We Know, What's Needed, and What's Next. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2):455:1–455:33, 2022.

[74] Raphael Serafini, Marco Gutfleisch, Stefan Albert Horstmann, and Alena Naiakshina. On the Recruitment of Company Developers for Security Studies: Results from a Qualitative Interview Study. In *Proceedings of the Nineteenth Symposium on Usable Privacy and Security*, pages 321–340, Anaheim, USA, 2023. USENIX Association.

[75] Raphael Serafini, Clemens Otto, Stefan Albert Horstmann, and Alena Naiakshina. ChatGPT-Resistant Screening Instrument for Identifying Non-Programmers. In *Proceedings of the IEEE/ACM 46th International Conference on Software Engineering*, ICSE '24. ACM, 2024.

[76] Hossein Shirazi, Bruhadeshwar Bezawada, Indrakshi Ray, and Charles Anderson. Adversarial Sampling Attacks Against Phishing Detection. In Simon N. Foley, editor, *Data and Applications Security and Privacy XXXIII*, pages 83–101, Cham, 2019. Springer International Publishing.

[77] Amit Kumar Sikder, Leonardo Babun, Z. Berkay Celik, Hidayet Aksu, Patrick McDaniel, Engin Kirda, and A. Selcuk Uluagac. Who's Controlling My Device? Multi-User Multi-Device-Aware Access Control System for Shared Smart Home Environment. *ACM Transactions on Internet of Things*, 3(4):27:1–27:39, 2022.

[78] Surya Srikar Sirigineedi, Jayesh Soni, and Himanshu Upadhyay. Learning-based models to detect runtime phishing activities using URLs. In *Proceedings of the 2020 4th International Conference on Compute and Data Analysis*, pages 102–106, New York, USA, 2020. ACM.

[79] Katta Spiel, Oliver L. Haimson, and Danielle Lottridge. How to do better with gender on surveys: a guide for HCI researchers. *Interactions*, 26(4):62–65, 2019.

[80] Andie Storozuk, Marilyn Ashley, Véronic Delage, and Erin A. Maloney. Got Bots? Practical Recommendations to Protect Online Survey Data from Bot Attacks. *The Quantitative Methods for Psychology*, 16(5):472–481, 2020.

[81] Mohammad Tahaei, Ruba Abu-Salma, and Awais Rashid. Stuck in the Permissions With You: Developer & End-User Perspectives on App Permissions & Their Privacy Ramifications. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, New York, USA, 2023. ACM.

[82] Mohammad Tahaei and Kami Vaniea. Recruiting Participants With Programming Skills: A Comparison of Four Crowdsourcing Platforms and a CS Student Mailing List. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, New York, USA, 2022. ACM.

[83] Jenny Tang, Eleanor Birrell, and Ada Lerner. Replication: How Well Do My Results Generalize Now? The External Validity of Online Privacy and Security Surveys. In *Eighteenth Symposium on Usable Privacy and Security*, pages 367–385, Boston, USA, 2022. USENIX Association.

[84] Christopher Thompson, Martin Shelton, Emily Stark, Maximilian Walker, Emily Schechter, and Adrienne Porter Felt. The Web's Identity Crisis: Understanding the Effectiveness of Website Identity Indicators. In *Proceedings of the 28th USENIX Security Symposium*, pages 1715–1732. USENIX Association, 2019.

[85] Ke Tian, Steve T. K. Jan, Hang Hu, Danfeng Yao, and Gang Wang. Needle in a Haystack: Tracking Down Elite Phishing Domains in the Wild. In *Proceedings of the Internet Measurement Conference 2018*, pages 429–442, New York, USA, 2018. ACM.

[86] Sarah Turner, Jason Nurse, and Shujun Li. When Googling It Doesn't Work: The Challenge of Finding Security Advice for Smart Home Devices. In Steven Furnell and Nathan Clarke, editors, *Human Aspects of Information Security and Assurance*, volume 613, pages 115–126, Cham, 2021. Springer International Publishing.

[87] Ishant Tyagi, Jatin Shad, Shubham Sharma, Siddharth Gaur, and Gagandeep Kaur. A novel machine learning approach to detect phishing websites. In *2018 5th international conference on signal processing and integrated networks*, pages 425–430, 2018.

[88] Kami E. Vaniea, Emilee Rader, and Rick Wash. Betrayed by updates: how negative experiences affect future security. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2671–2674, New York, USA, 2014. ACM.

[89] Francesco Vitale, Joanna McGrenere, Aurélien Tabard, Michel Beaudouin-Lafon, and Wendy E. Mackay. High Costs and Small Benefits: A Field Study of How Users Experience Operating System Upgrades. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 4242–4253, New York, USA, 2017. ACM.

[90] Grega Vrbančič, Iztok Fister, and Vili Podgorelec. Swarm Intelligence Approaches for Parameter Setting of Deep Learning Neural Network: Case Study on Phishing Websites Classification. In *Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics*, Novi Sad Serbia, 2018. ACM.

[91] Rick Wash. Folk models of home computer security. In *Proceedings of the Sixth Symposium on Usable Privacy and Security*, Redmond, USA, 2010. ACM.

[92] Rick Wash and Emilee Rader. Too Much Knowledge? Security Beliefs and Protective Behaviors Among United States Internet Users. In *Proceedings of the Eleventh Symposium On Usable Privacy and Security*, pages 309–325. USENIX Association, 2015.

[93] Rick Wash, Emilee Rader, Kami Vaniea, and Michelle Rizor. Out of the Loop: How Automated Software Updates Cause Unintended Security Consequences. In

10th Symposium On Usable Privacy and Security, pages 89–104. USENIX Association, 2014.

[94] Alma Whitten and J. D. Tygar. Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0. In *Proceedings of the 8th USENIX Security Symposium*, pages 169–184. USENIX Association, 1999.

[95] Daniel W. Woods and Sezaneh Seymour. Evidence-based cybersecurity policy? A meta-review of security control effectiveness. *Journal of Cyber Policy*, 8(3):365–383, 2023.

[96] Min Wu, Robert C. Miller, and Simson L. Garfinkel. Do security toolbars actually prevent phishing attacks? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 601–610, Montréal Québec Canada, 2006. ACM.

[97] Leon Würsching, Florentin Putz, Steffen Haesler, and Matthias Hollick. FIDO2 the Rescue? Platform vs. Roaming Authentication on Smartphones. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, New York, USA, 2023. ACM.

Appendix

The appendix contains the full demographic information from our re-replication study in Table 3. We also provide additional descriptive statistics on experts' and non-experts' advice ratings in Table 4.

To facilitate comparison between the current re-replication and the earlier studies from 2014 and 2018, we also provide statistical test results in Tables 5 and 6. Since we did not have hypotheses regarding comparisons of specific behaviors or pieces of advice, we used corrections for multiple testing on the reported p-values, specifically the Holm correction. We deviated somewhat from prior work in our choice of tests but provide effect sizes to enable comparison across publications. Since the assumptions for the χ^2 test were not met in eight cases, we chose to consistently use the Fisher's Exact test. Instead of using the Wilcoxon ranksum test which is intended for independent samples, we use the Wilcoxon signed rank test for dependent samples, as effectiveness and realism ratings came from the same participants and these data points cannot be considered independent. We report effect sizes (Cramer's V for Fisher's Exact tests and rankbiserial r (r_{rb}) for Wilcoxon signed rank tests to enable comparison with prior work beyond statistical significance.

		NE	E
Gender	Women	50.10 %	13.33 %
	Men	48.38 %	84.00 %
	Non-binary	1.01 %	0.00 %
	Self-described	0.10 %	0.00 %
	No answer	0.40 %	2.67 %
Age	18 - 24	11.31 %	22.67 %
	25 - 34	17.27 %	54.67 %
	35 - 44	17.47 %	16.00 %
	45 - 54	16.06 %	04.00 %
	55 - 64	25.96 %	1.33 %
	65 or older	11.82 %	1.33 %
	No answer	0.10 %	0.00 %
Education	Profession. Doctorate	1.62 %	1.33 %
	Doctoral Degree	1.72 %	5.33 %
	Master	15.25 %	56.00 %
	Bachelor	38.38 %	22.67 %
	Associates Degree	9.70 %	10.67 %
	Some college, no degree	18.08 %	0.00 %
	Tech./Trade School	3.33 %	4.00 %
	Other	11.41 %	0.00 %
	No answer	0.51 %	0.00 %
Occupation	Employed full-time	47.07 %	
	Employed part-time	13.94 %	
	Self-employed	11.01 %	
	Care-provider	0.30 %	
	Homemaker	4.75 %	
	Retired	11.41 %	
	Student	2.32 %	
	Looking for work	6.67 %	
	Other	1.92 %	
	No answer	0.61 %	
Work place	Industry		37.33 %
	University		42.67 %
	Corp. research lab		5.33 %
	Government		8.00 %
	Self-employed		5.33 %
	Other		1.33 %
Security experience	1 - 4.5 years		28.00 %
	5 - 9.5 years		45.33 %
	10 - 14.5 years		13.33 %
	15+ years		13.33 %
Location	US	100.00 %	5.33 %
	Europe	0.00	92.00 %
	Africa	0.00	2.67 %

Table 3: Full Demographic information for expert (E; $N = 75$) and non-expert (NE; $N = 994$) survey from 2024 sample

Advice questions	Expert – Goodness						Non-Expert – Effectiveness						Non-Expert – Realism					
	1	2	3	4	5	IDK	1	2	3	4	5	IDK	1	2	3	4	5	IDK
Use anti-virus	17	15	20	25	23	0	2	4	15	28	50	1	2	5	9	20	63	1
Install latest OS updates	0	4	12	27	57	0	1	4	13	30	51	1	1	3	9	23	63	0
Automatic updates	0	3	8	28	61	0	3	6	18	29	42	2	3	5	10	25	56	1
Update applications	0	8	23	35	35	0	1	5	16	33	45	1	1	3	9	27	58	1
Clear cookies	25	37	19	8	11	0	3	11	22	26	34	3	3	6	15	23	52	0
Check app permissions	9	25	31	23	12	0	1	5	17	28	48	2	1	4	12	26	56	1
Use unique pws	1	8	13	23	55	0	2	3	8	20	67	0	4	5	12	19	60	0
Use hard to guess pws	4	5	16	20	55	0	1	2	4	16	78	0	2	2	6	19	72	0
Don't write down pws on paper	12	20	24	13	28	3	14	12	16	15	41	3	19	10	12	13	44	2
Save pws locally	45	32	11	9	1	1	38	19	17	11	11	3	35	14	16	13	20	3
Use pw manager	1	1	13	28	56	0	7	8	20	26	35	3	11	9	17	19	41	3
Write down pws on paper	28	17	31	13	5	5	36	15	17	13	16	4	37	11	15	13	21	4
Check if HTTPS	5	9	16	32	37	0	2	4	11	27	55	2	2	3	11	20	64	1
Be careful in general	1	11	27	23	39	0	0	3	10	21	66	0	1	2	9	19	68	0
Be careful with links in e-mails	0	7	8	27	59	0	1	2	4	13	81	0	1	2	5	13	81	0
Visit only known websites	16	27	24	26	7	0	2	5	20	28	45	1	6	8	19	25	41	0
Use 2FA	3	8	16	29	44	0	1	2	5	17	74	1	2	3	9	21	66	0
Don't share private info	3	17	24	23	32	1	1	1	5	16	77	0	1	1	5	15	78	0
Use hardware token	11	25	27	17	17	3	5	4	19	23	33	16	11	11	23	21	23	10
Use authenticator app	0	8	20	40	29	3	2	3	11	25	50	8	5	6	15	22	47	4
Use passkeys	3	13	17	27	24	16	5	5	16	27	37	11	9	8	19	22	36	6
Use biometrics	4	8	20	29	35	4	4	3	11	25	55	4	8	6	11	18	55	3
Don't click on links from unknown	0	16	23	28	33	0	1	2	5	16	76	0	1	2	5	13	79	0
Don't enter pws on links from e-mail	0	7	16	25	52	0	1	2	3	14	79	1	1	1	4	14	80	0
Check URL	1	8	15	37	39	0	1	2	5	18	74	0	1	2	8	15	73	0
Don't open e-mail attachments from unknown	0	4	16	27	53	0	0	1	5	13	81	0	1	2	5	13	79	0
Use AI security advice	20	29	23	20	5	3	21	17	28	15	12	7	25	13	24	15	16	6

Table 4: Expert and non-expert ratings of advice. Differences of expert "goodness" ratings from non-expert realism and effectiveness ratings are marked in green, if the difference is less than 5 percent points and in orange, if the difference is more than 10 percent points per individual response category. We abbreviate "password" as **pw**.

Behavior	$N_{Experts}$	$N_{Non-Experts}$	Cramer's V	95% CI	raw p	adj. p
Do OS update	63	939	0.12	[0.01,1]	0.003	0.023
Use anti-virus	61	882	0.22	[0.17,1]	<0.001	<0.001
Remember passwords	75	990	0.17	[0.108,1]	<0.001	<0.001
Write down passwords on paper	75	990	0.14	[0.073,1]	<0.001	<0.001
Save passwords locally	75	990	0.14	[0.073,1]	<0.001	<0.001
Use password manager	75	990	0.21	[0.154,1]	<0.001	<0.001
Re-use passwords	75	990	0.13	[0.059,1]	<0.001	<0.001
Use 2FA	74	974	0.06	[0,1]	0.028	0.181
Use hardware token	70	924	0.20	[0.143,1]	<0.001	<0.001
Use authenticator app	72	930	0.20	[0.142,1]	<0.001	<0.001
Use passkeys	71	933	0.00	[0,1]	0.802	1
Use biometric auth	68	985	0.00	[0,1]	0.425	1
Check app permissions	68	968	0.03	[0,1]	0.226	0.903
Check URL bar	69	989	0.07	[0,1]	0.036	0.181
Check Https	54	973	0.22	[0.159,1]	<0.001	<0.001
Visit unknown websites	75	985	0.22	[0.161,1]	<0.001	<0.001
Enter password on link from e-mail	59	956	0.00	[0,1]	0.668	1
Open e-mail from unknown	71	986	0.30	[0.246,1]	<0.001	<0.001
Click on link from unknown	72	988	0.07	[0,1]	0.026	0.181
Share unnecessary personal info	73	986	0.21	[0.148,1]	<0.001	<0.001

Table 5: Fisher's Exact tests comparing expert and non-expert security behavior from the 2024 sample. We used the Holm correction to adjust p-values for multiple testing.

Advice	effectiveness		realism		V	r_{rb}	95% CI	raw p	adj. p
	M	SD	M	SD					
Use anti-virus	4.24	0.95	4.40	0.97	18054	-0.39	[-0.448,-0.325]	<0.001	<0.001
OS updates	4.27	0.90	4.45	0.86	16566	-0.43	[-0.489,-0.371]	<0.001	<0.001
Automatic updates	4.02	1.08	4.27	1.04	18984	-0.49	[-0.542,-0.432]	<0.001	<0.001
Update applications	4.16	0.95	4.40	0.86	17799	-0.52	[-0.567,-0.461]	<0.001	<0.001
Clear cookies	3.79	1.15	4.15	1.08	24534	-0.54	[-0.585,-0.481]	<0.001	<0.001
Check app permissions	4.19	0.95	4.33	0.92	24802	-0.28	[-0.343,-0.209]	<0.001	<0.001
Use unique pws	4.49	0.88	4.27	1.10	41798	0.42	[0.356,0.475]	<0.001	<0.001
Use strong pws	4.68	0.70	4.58	0.80	19986	0.29	[0.227,0.359]	<0.001	<0.001
Don't write down pws on paper	3.58	1.48	3.54	1.58	46832	0.04	[-0.029,0.117]	0.419	1
Save pws locally	2.37	1.40	2.68	1.55	35896	-0.35	[-0.409,-0.279]	<0.001	<0.001
Use pw manager	3.78	1.23	3.73	1.39	44956	0.08	[0.009,0.155]	0.13	1
Write down pws on paper	2.59	1.50	2.69	1.60	26445	-0.17	[-0.239,-0.095]	0.005	0.056
Check if https	4.32	0.94	4.43	0.90	16520	-0.31	[-0.37,-0.238]	<0.001	<0.001
Be careful in general	4.49	0.81	4.52	0.83	12146	-0.11	[-0.179,-0.037]	0.118	1
Be careful with links in e-mails	4.71	0.70	4.71	0.67	6330	0.00	[-0.077,0.067]	0.956	1
Visit only known websites	4.08	1.02	3.89	1.20	50616	0.35	[0.283,0.41]	<0.001	<0.001
Use 2FA	4.62	0.76	4.47	0.88	22838	0.46	[0.397,0.511]	<0.001	<0.001
Don't share private info	4.69	0.67	4.68	0.72	8094	0.04	[-0.033,0.111]	0.633	1
Use hardware token	3.89	1.16	3.43	1.32	60442	0.55	[0.496,0.606]	<0.001	<0.001
Use authenticator app	4.29	0.96	4.10	1.14	34700	0.38	[0.311,0.439]	<0.001	<0.001
Use passkeys	3.98	1.11	3.81	1.27	38542	0.28	[0.212,0.353]	<0.001	<0.001
Use biometrics	4.28	1.03	4.13	1.24	26838	0.31	[0.24,0.373]	<0.001	<0.001
Don't click on links from unknown	4.66	0.72	4.68	0.73	8733	-0.10	[-0.166,-0.024]	0.218	1
Don't enter pws on links from e-mail	4.71	0.69	4.71	0.66	5548	-0.05	[-0.117,0.026]	0.602	1
Check URL bar	4.63	0.72	4.58	0.81	14439	0.20	[0.129,0.267]	0.006	0.061
Don't open email attachments from unknown	4.73	0.64	4.67	0.74	8390	0.24	[0.171,0.307]	0.005	0.056
Use AI advice	2.79	1.30	2.83	1.42	23319	-0.07	[-0.149,0]	0.222	1

Table 6: Wilcoxon signed rank tests comparing non-expert effectiveness and realism ratings of security advice from the 2024 sample. We abbreviate "password" as **pw**. We used the Holm correction to adjust p-values for multiple testing.