Prototyping Usable Privacy and Security Systems: Insights from Experts

In International Journal of Human–Computer Interaction (HIHC)

Florian Mathis, University of Glasgow, United Kingdom

florian.mathis@glasgow.ac.uk

Kami Vaniea, University of Edinburgh, United Kingdom

kvaniea@inf.ed.ac.uk

Mohamed Khamis, University of Glasgow, United Kingdom

mohamed.khamis@glasgow.ac.uk

Prototyping Usable Privacy and Security Systems: Insights from Experts

Florian Mathis[a,b,*], Kami Vaniea[b] and Mohamed Khamis[a]

[a]University of Glasgow, Scotland, United Kingdom; [b]University of Edinburgh, Scotland, United Kingdom

ABSTRACT
Iterative design, implementation, and evaluation of prototype systems is a common approach in Human-Computer Interaction (HCI) and Usable Privacy and Security (USEC); however, research involving physical prototypes can be particularly challenging. We report on twelve interviews with established and nascent USEC researchers who prototype security and privacy-protecting systems and have published work in top-tier venues. Our interviewees range from professors to senior PhD candidates, and researchers from industry. We discussed their experiences conducting USEC research that involves prototyping, opinions on the challenges involved, and the ecological validity issues surrounding current evaluation approaches. We identify the challenges faced by researchers in this area such as the high costs of conducting field studies when evaluating hardware prototypes, the scarcity of open-source material, and the resistance to novel prototypes. We conclude with a discussion of how the USEC community currently supports researchers in overcoming these challenges and places to potentially improve support.

1. Introduction

Prototyping is an integral part of human-centered research and design (Fallman, 2003; Ogunyemi, Lamas, Lárusdóttir, & Loizides, 2019; Wobbrock & Kientz, 2016). Wobbrock and Kientz (2016) argue that one of the main types of research contributions in Human-computer Interaction (HCI) is artifact contributions: where researchers design inventive prototypes, such as new systems, tools and techniques that demonstrate novel forward-looking possibilities, or generate new insights through implementing and evaluating the prototypes (e.g., (Baudisch, Sinclair, & Wilson, 2006; Greenberg & Fitchett, 2001; Ishii & Ullmer, 1998; Lopes, You, Cheng, Marwecki, & Baudisch, 2017; Lopes, You, Ion, & Baudisch, 2018)). Usable Privacy and Security (USEC) research is not an exception. USEC researchers have brought forth a plethora of novel usable privacy and security systems that extended state-of-the-art and facilitated new insights (e.g.,

---

*Corresponding Author: Florian Mathis

Email addresses: florian.mathis@glasgow.ac.uk (Florian Mathis), kvaniea@inf.ed.ac.uk (Kami Vaniea), mohamed.khamis@glasgow.ac.uk (Mohamed Khamis)

(De Luca et al., 2014; De Luca, von Zezschwitz, Nguyen, et al., 2013; Hayashi, Riva, Strauss, Brush, & Schechter, 2012; Krombholz, Hupperich, & Holz, 2016)) – some of which found their way to wider adoption, such as PassPoints, Pass-Go and DAS which inspired Android's lock patterns (Jermyn, Mayer, Monrose, Reiter, & Rubin, 1999; Tao & Adams, 2008; Wiedenbeck, Waters, Birget, Brodskiy, & Memon, 2005). At the same time, USEC researchers have argued for the importance of human-centered design since the 1970's, when Saltzer and Schroeder (1975) outlined that security protection mechanisms require "psychological acceptability". This position was taken further by researchers from both the security and HCI communities (Adams & Sasse, 1999; Whitten & Tygar, 1999; Zurko & Simon, 1996).

Conducting research that involves prototyping comes with unique challenges, such as hardware deployments in ecologically valid contexts and evaluations with adequate sample sizes, that hinder its undertaking. Our work provides: 1) the first interview-based insight into the challenges faced by USEC experts when designing, implementing, evaluating and also publicizing research that is based on prototyping usable privacy and security systems, and 2) ways forward to support research in this direction on both the individual researcher and the wider community level. We interviewed twelve expert and nascent USEC researchers from academia and industry who have made significant contributions to USEC research and whose work involved prototyping novel systems to unveil and better understand their research challenges. Our interviewees include full/associate/assistant professors, researchers from large tech companies, consultants, and senior PhD candidates. Our work tackles the following two research questions:

- **RQ$_1$:** Where, if any, are the bottlenecks USEC experts face when designing, implementing, and evaluating usable privacy and security prototype systems?

- **RQ$_2$:** What does the USEC community need to better facilitate the transition of artifact contributions into practice?

We present 9 key challenges impeding artifact contributions in USEC, including challenges that have not seen in-depth discussion in prior literature, e.g., the implementation challenges due to scarcity of open-source material; difficulties conducting ecologically valid studies, especially when evaluating hardware usable privacy and security solutions; and the lack of publication venues where novel evaluated USEC systems are encouraged. We propose five ways the USEC community can support overcoming these challenges, such as encouraging collaborations between academia, industry and across research groups, being open to novel evaluated solutions, and encouraging development of new methodologies to cope with high costs of ecologically valid field studies and the shortcomings of lab studies.

We aim to raise awareness of the existing challenges and start a critical discussion and self-reflection on how the USEC community operates, provoke change in how the community addresses work that involves prototyping USEC systems, and discuss our experts' voiced challenges in the light of neighboring communities such as HCI, Mobile HCI, Ubicomp. The insights from USEC experts coupled with the in-depth discussions presented in this work should be valuable to the USEC community as well as neighboring communities.

2. Background & Related Work

The term "usable privacy and security research" refers to research that touches both on human-factors work such as human-computer interaction, design, and user experience as well as on privacy and security issues such as user authentication, email security, anti-phishing, web privacy, mobile security/privacy, and social media privacy. Because the research is by its nature interdisciplinary, it inherits the research approaches and challenges from all areas it touches on. For example, many of the research methodologies are drawn from the HCI community (Garfinkel & Lipford, 2014) which has a rich history in user-based research. Yet, many approaches need to be adapted to handle the sensitive nature of security and privacy work. For example, finding ways to study ATM PIN entry in a way that does not break laws, endanger participants, or leak sensitive data while also ensuring the evaluation is ecologically valid are all challenging (De Luca, Langheinrich, & Hussmann, 2010; Volkamer, Gutmann, Renaud, Gerber, & Mayer, 2018). As a result, privacy and security researchers struggle with getting access to "real" user data or need to spend significant additional effort. In this paper, we aim to identify the set of challenges that are particularly problematic to the subset of the USEC community that conducts prototyping-related research.

2.1. USEC Research and its Challenges

Past efforts have organized existing research in particular domains within USEC. Iachello and Hong (2007) outlined approaches, results, and trends in research on privacy in HCI. In their work, they analyzed academic and industrial literature published between 1977 and 2007. They described some legal foundations and historical aspects of privacy, which included designing, implementing, and evaluating privacy-affecting systems. Work by Acar, Fahl, and Mazurek (2016) reviewed state-of-the-art USEC research that typically focused on end-users, and laid out an agenda to support software developers. A more general review of USEC research by Garfinkel and Lipford (2014) covered the state of USEC research in 2014, and suggested future research directions. Their work highlighted that *"only by simultaneously addressing both usability and security concerns will we be able to build systems that are truly secure"* (Garfinkel & Lipford, 2014, p.vi), emphasizing the need for novel well-evaluated systems that address both security and usability from the beginning of the design process.

Alt and von Zezschwitz (2019) highlighted the need to develop study paradigms for collecting data while minimizing the required effort for both participants and researchers. Alt and von Zezschwitz (2019) and Bianchi and Oakley (2016) also underlined that the fast pace of emerging technologies (e.g., wearables such as smart glasses and smart watches) comes with novel challenges that require rapid adaptions of user-centered usability, security, and privacy research. For example, while wearable devices can enable novel authentication techniques (e.g., (Chun et al., 2016; Liu, Cornelius, Rawassizadeh, Peterson, & Kotz, 2018), they also require researchers to adjust their research methods to consider new contexts and privacy implications (Bianchi & Oakley, 2016). The book by Garfinkel and Lipford (2014) dedicates three pages (p. 4-6) in the intro to discussing challenges that make USEC research hard, including challenges around conducting ecologically valid studies due to priming caused by the study itself and challenges around designing in the presence of an adversary that is actively attempting to attack the user. USEC research has also been previously found to be challenging because security is often not users' primary task (A. M. Sasse, Brostoff, &

Weirich, 2001). Studying sensitive and private contexts often also requires additional effort and resources due to ethical and legal constraints, as highlighted in prior work (De Luca et al., 2010; Trowbridge, Sharevski, & Westbrook, 2018).

While the works above provide valuable observations, most of them are drawn either from a single study or a review of written works. Their focus is also often on capturing the current state of the field or providing structured book-like observations for students. Work that captures common opinions and "hallway chatter" among USEC researchers that is typically less structured are more rare. One of the few works in this direction is a work by Sasse et al. M. A. Sasse, Smith, Herley, Lipford, and Vaniea (2016) which reports on a record of a conversation among experts about the usability-security trade-off. These works are valuable because they capture the attitudes of practitioners in the field in a more candid way. Our work aims to provide this kind of "hallway knowledge" view of the challenges faced by USEC researchers who design, develop, and evaluate prototypes.

2.2. A Glance at USEC's Prototyping Challenges

Building and testing prototypes is a common approach in USEC when the specifics of a design are likely to impact how users interact with it. Prototyping is commonly used in areas like authentication where the physical design of the input mechanism can have a large impact on a user's input speed and accuracy as well as an attacker's ability to remotely view and replicate their actions. We use the range of USEC's prototypes that we discuss below to provide readers with an idea about the available prototype systems in USEC research. Note that people often have different expectations of what a prototype is. As a result, finding one definition that covers all different research fields and prototype variants is challenging.

> "Everyone has a different expectation of what a prototype is. [...] Is a brick a prototype? The answer depends on how it is used. If it is used to represent the weight and scale of some future artifact, then it certainly is: it prototypes the weight and scale of the artifact." - (Houde & Hill, 1997, p.368)

While we are not aware of an existing overview of prototyping challenges in USEC, several researchers have commented on specific challenges they faced when designing and testing USEC-related prototypes. Reilly et al. (2014) presented a software toolkit for USEC research in mixed reality collaborative environments and emphasized that complex prototypes can be difficult to set up correctly. They also argued that their toolkit was limited by the functionalities of the base platform they used (i.e., Open Wonderland[1]). Zeng and Roesner (2019) built a prototype smart home app to find answers to how a smart home should be designed to address multi-user security and privacy challenges and what security and privacy behaviors smart home users exhibit in practice. In their work, they identified challenges introduced by their used Smart-Things API. Activity notifications could not be used to attribute changes in the home state to particular third-party apps. Other limitations were introduced by the underlying operating system. Persistent low priority notifications were only implemented on Android but not on iOS as the notification center did not support these notifications (Zeng & Roesner, 2019). Other works reported that their prototype limitations resulted in lower ecological validity. Hundlani, Chiasson, and Hamid (2017), for example, stated that their prototypes were created for research purposes only and were not

---

[1]Open Wonderland enables researchers to build interactive and multi-user virtual worlds (http://www.openwonderland.org/, accessed 08/03/2021)

on the level of finished products.

When it comes to hardware prototypes, USEC researchers have reported a variety of additional challenges. Physical prototypes are often made in research labs and therefore are physical approximations rather than professionally designed products, which can lead to confounds around usability. For example, using two connected mobile phones to provide users with a back and front display for user authentication enabled testing of the idea, but at the same time negatively impacted users' authentication experience due to the prototype's weight (De Luca, von Zezschwitz, Nguyen, et al., 2013). The work by Chen et al. (2020) showed that achieving a form factor similar to the original product can be technically challenging. Their wearable jammer to protect users' privacy was indeed larger than a typical bracelet. Other work reported that the form factor of their privacy-protecting prototype was not perceived well by users (Perez, Zeadally, Matos Garcia, Mouloud, & Griffith, 2018). Prototypes are also often built using existing hardware and software which can limit the range of what they can accurately do. Mhaidli, Venkatesh, Zou, and Schaub (2020), for example, built a smart speaker prototype but encountered robustness issues with their Kinect camera when tracking users' eye movements. In a similar vein, Schaub et al. (2014) faced reliability issues with their presence detection and identification system, which likely resulted in more conservative privacy settings than preferred by their participants.

The challenges mentioned above are likely only a small percentage of the types of problems USEC researchers face when designing, developing, and evaluating prototypes. We aim to expand on these existing observations about prototyping challenges by talking with researchers about their experiences and challenges that may be well known in the community, but are not necessarily reported in publications.

2.3. Contribution Over Prior Work

In this work we present an overview of challenges faced by experts who design, develop, and evaluate prototypes as part of their USEC-related research. While other works have touched on what makes USEC research challenging in general, and individual works have commented on challenges they have faced in completing their research, there has not yet been a structured attempt to elicit challenges experienced by USEC researchers that use prototypes in their work. In this work we put forward such a compilation of experienced challenges.

Such a compilation is valuable and novel, and it also sits within the wider contexts of the HCI, security, and privacy fields. Some of our identified challenges have been identified previously, for example, the challenge of finding research participants is well known in HCI. However, we argue that there is value in compiling the set of challenges that most impact USEC researchers and to put these challenges into a USEC context. For example, USEC makes heavy use of deception studies where the participant is told that the study is about, say, testing a social networking site, but the research is actually about authentication or privacy. The use of deception makes it impossible to re-use participants, so the well known HCI challenge of finding participants has a particular shape in USEC work. In this paper we present the challenges faced by our participants and how those challenges manifested in their research.

5

3. Methodology

This section describes our recruiting process, the structure of our interviews, our research approach and analysis, and some potential limitations of our work.

3.1. Recruiting USEC Experts

We completed an ethical review through the University of Glasgow College of Science & Engineering Ethics Committee in advance of participant recruitment. Potential interviewees were selected with the goal of obtaining a mix of researchers and practitioners who are experienced in USEC; thus, work at the intersection of HCI and security & privacy research. We sought those who both published works and had hands-on experience in the design, implementation, and evaluation of usable privacy and security systems. To compose a list of potential interviewees, we started with a rough literature review to identify authors and then added people we knew about already in the area to fill out the list. We searched the ACM Digital Library, IEEE Xplore, and Google Scholar to find scholars with published USEC work at highly ranked HCI and security venues (e.g., ACM CHI, USENIX SOUPS, IEEE S&P). We started with broad search terms like *"usable security"*, *"usable privacy"* that formed the basis of our search and then followed-up with more specific search terms that are relevant for our research: *"security prototype"*, *"privacy prototype"*. We then reviewed papers to identify those that included building a security and/or privacy-protection solution and a user-centered evaluation. To further improve our coverage, we also used a snowball approach: the references in the papers were reviewed for relevant titles and added to the list of reviewed publications. We used Google Scholar and dblp to determine the publication profile and experience of the identified authors in the area. Relevant identified publications were recorded for later use in the interviews. From this process we identified a pool of 56 potential interviewees who have significant expertise in usable privacy and security and prototyping. We sorted the list with an eye towards multiple variables: selecting people with a range of seniority, university, industry, country, research domain, and experience publishing systems solution papers in USEC venues. Researchers who were more senior and had recent USEC prototype publications were ranked higher.

We then sent invitations (see template in Appendix A) asking if the person was willing to be interviewed about their research. Although recruiting senior people is time-consuming and challenging, we were able to secure fourteen responses (70%) from twenty invitations. Two then declined due to unavailability, the remaining twelve agreed to participate. Eleven interviews took place via Skype and were audio and video recorded with consent. One preferred an email interview, which is a viable alternative (Meho, 2006). As we progressed through the interviews, few novel insights emerged after the tenth interview. We continued with two more interviews and observed nothing new in the twelfth (theoretical saturation) (Guest, Bunce, & Johnson, 2006). We, therefore, decided not to send out additional interview requests.

Demographics of the USEC Experts and Interview Material. Our final sample had 12 USEC experts (4 females, 8 males). Our interviewees are from the US, Europe, and Asia, and work in academia (6), industry (2), or in both academia and industry (4). At the time of the interviews, 10 interviewees held a PhD (1 full professor / 4 associate professors / 1 assistant professor / 1 adjunct professor & security

Table 1. Our interviewees have published a significant number of work ($\overline{x}_{\mathrm{pub}}$=123.42) that is highly cited ($\overline{x}_{\mathrm{cite}}$=3740.75). Note that the data reported is from early 2020.

Anonymized Participants*	Publications	Citations	h-index	Job title	Academia	Industry
P1	[0,50]	[0,100]	[0,5]	PhD candidate	✓	✗
P2	[50,100]	[2.500,5.000]	>30	User Experience Researcher	✗	✓
P3	[100,250]	[2.500,5.000]	>30	Associate Professor	✓	✗
P4	[0,50]	[100,250]	>5	PhD candidate & UX Researcher	✓	✓
P5	[0,50]	[100,250]	>5	USEC Research Engineer	✗	✓
P6	[50,100]	[500,1.000]	>10	Assoc. Prof. & UX Researcher	✓	✓
P7	[0,50]	[0,100]	>5	Research Fellow in USEC	✓	✗
P8	[250,500]	[10.000,20.000]	>50	Full Professor	✓	✓
P9	[100,250]	[2.500,5.000]	>25	Associate Professor	✓	✗
P10	[50,100]	[1.000,2.500]	>15	Associate Professor	✓	✗
P11	[0,50]	[1.000,2.500]	>10	Assistant Professor	✓	✗
P12	[250,500]	[10.000,20.000]	>50	Adj. Prof. & Security Research Scientist	✓	✓
$\overline{x}$	123.42	3740.75	22.5	-	$\sum$ 10	$\sum$ 6

*To protect experts' identities, we mention here intervals for the number of publications, citations, and h-index.

research scientist / 1 user experience researcher / 1 USEC research engineer / 1 research fellow). We also included two senior PhD candidates who had published usable privacy and security research in top-tier venues and received best paper awards. Their inclusion widened the covered spectrum as they had more recent hands-on experience in implementing systems and conducting user-centered evaluations. All interviewees worked in the broader field of usable privacy and security including, but not limited to, user authentication, anti-phishing efforts, mobile security and privacy, and web privacy. Our interviewed experts have on average 123.42 publications *(max=386, min=18, SD=129.81)*, 3740.75 citations *(max=14627, min=25, SD=4857.28)*, and an h-index of 22.5 *(max=56, min=3, SD=16.69)*. All reported numbers (i.e., publications, citations, h-index) involve all kinds of publications, including usable privacy and security works. We report the overall numbers because all publications eventually contribute to a researcher's h-index, and extracting the number of USEC-specific papers in a precise way is challenging. The final set of publications (N=27) used to setup context during interviews ranged from 2010 to 2019 *(Md=2018)*. Out of the 27 publications used in the interviews, 14 papers comprise software-based prototype systems and 9 comprise hardware components. We also used four additional USEC papers from experts we interviewed, three of which are considered to be highly influential in USEC and the fourth one reports on research on an in-the-wild deployed security system. One of these additional publications discussed, for example, the last decade of usable security prototype systems and outlined learned lessons when developing and evaluating USEC prototypes. Table 1 shows an overview of our participants in an anonymised form.

3.2. Interview Structure

We conducted semi-structured interviews informed by the content of the interviewees' publications which the interviewer familiarized themselves with in advance. All publications were drawn from the initial literature review that we used for our sampling procedure, outlined in Section 3.1. The corresponding publications were then used as example papers and we attached them to the initial email request (see Figure A1 in Appendix A). This allowed us to efficiently use the interviewees' time and add context to their opinions. This also facilitated detailed discussions, allowing the interviewee to explore examples and the interviewer to ask informed follow up questions. We chose semi-structured interviews because they allowed us to ask the same high-level questions to all participants, and also ask follow up questions and encourage participants to discuss their past experiences. All interviews covered the following questions, and

follow up questions were prepared and used if needed (see full list in Appendix B):

Typical Research Journey from Idea to Publication. The purpose of this question was to understand how the interviewee normally progresses from a research idea to publication(s), and how that progression occurs within their broader research community. Journeys typically included topics such as: idea generation, resources, prototype development, idea refinement, evaluation, publication.

Research Challenges and Limitations. We asked our interviewees about the challenges they faced in conducting research that involved implemented solutions and their opinion about the more general challenges and limitations of USEC research that includes prototyping.

The Ecological Validity of Current Evaluations. We collected insights on different study types that our USEC experts employed. We also aimed to understand obstacles to conducting fully ecological valid evaluations. Inspired by recent work that argued for developing novel methodologies to understand and design for emerging technologies and mitigate new threats (Alt & von Zezschwitz, 2019; Garfinkel & Lipford, 2014), we asked our experts whether or not they see the current evaluation approaches of the USEC community as the way to go in the future or if they would prefer to see alternative evaluation approaches.

Finally, we debriefed our interviewees, asked them if they have any final questions or thoughts, and concluded with an informal chat. Interviews lasted 48.5 minutes on average. We offered all interviewees an £8 online shopping voucher. Some of them waived the compensation due to different reasons (e.g., donate it, keep it for other research projects).

3.3. Research Approach and Data Analysis

We applied open coding followed by thematic analysis inspired by Grounded Theory (Corbin & Strauss, 2014) on our interview data. We decided to a) apply open coding to build the insights and key challenges directly from the raw data of our expert interviews, and b) use a thematic analysis inspired by Grounded Theory (Corbin & Strauss, 2014) to uncover the main concerns and challenges of experts in the field of USEC when prototyping usable privacy and security systems. We also conducted an initial literature review prior to the interviews to better understand the research area and line up potential interviewees (as previously described in Section 3.1), who we then contacted by email. Doing this enabled us to familiarize ourselves with the expert's works and access a promising USEC sample for our investigation and research questions (Hoda, Noble, & Marshall, 2011). For the data collection, we conducted semi-structured interviews with open-ended questions. While interviews were ongoing, two authors regularly met to discuss 1) the notes taken by the interviewer about interesting observations and thoughts that emerged during the interviews and 2) the publications associated with upcoming interviews. These meetings allowed the researchers to regularly reflect on the findings and keep those points in mind in further interviews. Once all the interviews were completed, the lead author transcribed all audio recordings and open coded the transcriptions. The initial open coding scope was drawn from the regular discussion meetings and the lead researcher also took Memos (Saldaña, 2015) when conducting the open coding. A second researcher went through the interview raw data and added additional Memos. This process generated 325 open codes and

93 memos. The lead author then organised all codes and printed those out to have a paper-based piece for each code. Two authors then conducted a paper-based affinity diagram of the open codes (Kawakita, 1991). The transcript, Memos, and audio were revisited when additional information about a code's context was needed. The authors organized the codes into groups which were then further refined into themes.

In summary, we went through an initial literature review to compose a list of potential interviewees and followed with semi-structured interviews to collect USEC experts' opinions and insights that we then transcribed and further analyzed (Hoda et al., 2011). We present the themes, experts' voiced comments, and the key challenges when prototyping usable privacy and security systems in Section 4 and tie our findings with previous literature in an in-depth discussion in Section 5.

3.4. Limitations

While the approaches we use are common in human-centered and usable privacy and security research, some of our specific decisions have limitations to keep in mind. First, we selected experienced researchers who have been successful in publishing works involving prototypes in USEC venues. Their experience is valuable, but it is also biased towards those who ultimately succeeded in publishing. The challenges faced by those who tried and failed to conduct this type of research due to issues such as lack of mentorship, or choosing too challenging of problems are therefore not well represented here. Additionally, we aimed to talk to USEC researchers who have been successful and have been through a range of failures, which is an accompanying element when being successful in academia[2]. Although our sample also includes more junior researchers (e.g., P1, P7), we encourage future work to look into a sample of junior researchers only and compare their thoughts and opinions on USEC's prototyping challenges to the ones reported in our work. Moreover, interviews with experts in USEC on a research and community level might not have captured all sides of the conversation; for example, the views of entities such as research institutions and funding agencies are not covered. We leave this to future work. Our participants were also likely biased by the publications we selected and sent to them in advance of the interview. Pre-selecting publications helped both the interviewer and the interviewees scope the interviews in a time-productive manner. But the scoping also likely had an impact on the topics interviewees chose to discuss. Four of the participants, two pairs, had co-authored papers in our reviewed paper set. Given the seniority of some participants and the size of the field, such a situation is expected. However, we were careful not to use the same publication in more than one interview session to ensure that experts' voiced comments do not revolve around the same publications. Finally, interviews were also retrospective in nature, focusing on past experience and opinions about the area. While retrospective interviews can be quite effective for learning about rare events or those that take place over a long time period, they also suffer from a bias towards memorable events. Our interviewees described projects where the initial idea generation was sometimes years in the past likely resulting in some issues of memory bias.

---

[2]Two exemplary blog posts about being successful in academia and its accompanied failures can be read here: https://www.universityaffairs.ca/career-advice/career-advice-article/redefining-success-and-failure-in-academia/, accessed 09/03/2021 or https://www.timeshighereducation.com/features/scholars-divulge-their-biggest-mistakes, accessed 09/03/2021

4. Results

Below, we present our key findings: 1) threat modeling, 2) prototyping USEC systems, 3) sample size and selection, 4) evaluations, 5) USEC's research culture, and 6) USEC's real-world impact. Protecting expert participant anonymity is challenging (Saunders, Kitzinger, & Kitzinger, 2015; Scott, 2005; van den Hoonaard, 2003), so we refer to interviewees using a participant number (P1 to P12) and use *they* for all participants. In advance of presenting each key challenge, we use short preambles to set the frame of the challenge and introduce readers to the topic.

4.1. Threat Modeling is not Straight Forward

Threat modeling is commonly used in privacy and security research to describe the assumed skills and capabilities of an attacker. Many input and feedback methods can be observed by bystanders, which led to a lot of emphasis on shoulder surfing in the past (Bošnjak & Brumen, 2020). Shoulder surfing attacks, for example, often assume that the attacker can get physically close to the user or has an observation device like a camera. The considered threat impacts the design and evaluation of usable privacy and security prototypes because researchers need to consider said threats in the design and development process. While there are many different threats including, for example, social engineering attacks (Krombholz, Hobel, Huber, & Weippl, 2015) and online/offline guessing attacks (Gong, Lomas, Needham, & Saltzer, 1993), shoulder surfing as a threat model as studied by, for example, Brudy, Ledo, Greenberg, and Butz (2014); De Luca, Denzel, and Hussmann (2009); George, Khamis, Buschek, and Hussmann (2019); Mathis, Williamson, Vaniea, and Khamis (2020), was particularly discussed by several experts who exhibited a range of opinions about what constitutes a "realistic" threat model. Many of the interviewed experts focused on authentication research in the past, which is not surprising given that authentication has been a major theme in USEC research with Garfinkel and Lipford (2014) spending 27 pages on the subject compared to 10 on phishing. Consequently, the example of shoulder surfing threat models was brought up several times in reference to how threat models, prototype systems, and study designs can interact. Our expert interviews revealed two opposing opinions regarding valid threat models that address security, which we discuss below.

P5 argued that the relevance of shoulder surfing attacks depends heavily on the context, and explained that the threat has different implications in different countries and that these attacks definitely scare them.

> "In the U.S. as well as in Europe you may not really feel [that] shoulder surfing attacks are something that you should really care about [...]. In over-populated countries like India you have a lot of people [...] when you go to an ATM machine or to places like coffee shops [...] there are like three people standing right behind you. [In these cases,] shoulder surfing is a really big problem." **- P5**

Taken together, P5 voiced that cultural differences in perceived personal space impact susceptibility to shoulder surfing (Remland, Jones, & Brinkman, 1995). In a simialr vein, P8 outlined that such attacks are actually realistic in the real world and further argued that researchers have to consider both, user concerns as well as the point of view of experts to accurately assess the value and validity of certain threats.

> "If you actually keep asking the users about what they are worried about; often they are less worried about the NSA and more worried about their parents/their partner." **- P8**

However, even if a threat model seems to be appropriate for a given context and is of particular importance to end-users, experts had different opinions on the value of specific threat models and some mentioned that shoulder surfing attacks are overrated and uninteresting.

> "Shoulder surfing is a problem, but it's hugely overblown." **- P9**

> "Fundamentally for me the problem with observation attacks is, [they] are not that interesting from a security point of view, it's a real niche attack [...] [researchers] report performance against observation attacks with a very narrow threat model: "can you see it"; which is incredibly, it's very very narrow." **- P3**

P2 emphasized the problem that there is no common agreement among USEC experts regarding the validity of specific threat models.

> "[researchers] think they use the worst case scenario, but actually they did not."**- P2**

P7 further described the threat modeling challenges using shoulder surfing as an example and emphasized that there is a clear mismatch between researchers' assumptions and the reality and that it is important to consider social norms when studying threat models because "people move closer than [researchers] actually thought they ever would, or they stay further away because they respect people's social norms"(**P7**).

KEY CHALLENGE #1

Experts' opinions regarding the value of specific threat models vary widely. A good threat model needs to match the contextual realities of users, but those realities are not always known or may only impact a specific subset of users, making threat modeling a non-trivial part of research.

4.2. Prototyping USEC Systems

Prototyping is an integral part of human-centred research and design (Fallman, 2003; Wobbrock & Kientz, 2016), one of the main types of research contributions in HCI research (Wobbrock & Kientz, 2016), as well as a major theme in USEC research (Garfinkel & Lipford, 2014). We observed themes around the hardware challenges when building usable privacy and security prototypes (4.2.1) and around the deployment and corresponding evaluation challenges (4.2.2).

4.2.1. Development & Hardware Challenges

Experts voiced that developing usable privacy and security prototype systems is challenging and costly, partially due to limited access to appropriate hardware and the limited prototyping expertise of USEC researchers.

> "I think we actually really need more collaborations between the usable security people and the people who are fairly close to building [hardware prototypes]." **- P8**

P1 voiced that they faced issues with the eye tracker due to, for example, inappropriate lighting conditions. They also reported that the interplay between multiple hardware components caused them some issues and that this resulted in significant more effort, additional pilot tests, and in excluding data from the actual user study.

> "I combined [the hardware] all together [...] and then [faced] issues [...] because they are all working with infrared and [operate] on the same wave length." **- P1**

> "If you recruit 50 participants [...] you have to discard five to ten participants because the eye tracker is not working." **- P1**

Our experts also voiced that these limitations lead to many prototype systems *"[that] were made very quickly [and] are not well made" (P3)* or that hardware is used inappropriately, threatening ecological validity.

> "We slapped the phone on [a user's] wrist and put a little active part in a corner, so it was sort of a like big wrist watch but it was not usable [...] the validity of using a phone on users' wrist is relatively low." **- P3**

Experts attested that the lack of appropriate hardware, partially due to a lack of funding, is a fundamental problem.

> "Usually we do not have funding to buy new equipment [...] then we have to come up with ideas of how we can build that hardware." **- P4**

P2 mentioned that such bottlenecks have a noticeable impact on USEC research with their prototype being significantly heavier than mobile devices at that time.

> "A lot of negative feedback in those evaluations was around the weight of the prototype [...] [the weight] made it more difficult to use [the prototype]." **- P2**

Some experts even mentioned that they had to adjust their research projects due to the lack of appropriate research equipment.

> "We try to have as fast as we can the first prototype and see what are the challenges from the development side because often we need to alter the project to fit to the equipment we have." **- P4**

Other experts further mentioned that setting up hardware components at their intended place can be challenging and that these physical restrictions often force them to come up with alternative solutions: *"I didn't really manage to put [the front camera] exactly in the middle because the eye tracker was [already] there" (P1).*

4.2.2. Deployment & Evaluation Challenges

When it comes to the evaluation of USEC prototypes and conducting research that goes beyond in-lab investigations, experts explained they had a hard time in evaluating their systems and that there are a lot of issues around deployability, especially in the case of using new hardware. Although USEC experts strive for real-world deployments to increase ecological validity, P4 still sees the transferability of findings to users' everyday life as one of the major problems.

> "The major problem with evaluating privacy and security systems is that how can you visualize that the users are acting the way they would act if they would [use] it in their everyday life." **- P4**

P2 further voiced that deploying their prototype to a large sample was impossible and explained the situation with having access to only one device.

> "There's a lot of issues around deployability, specifically when it comes to using new hardware [...] the deployment was impossible [...] we had one device and that device we could hand out to one person at a time." **- P2**

When using new hardware, our interviewees also highlighted that the cost of failure might be high and that it is important to invest only in equipment that is likely to become publicly available in the future and provides promising future use cases. P2

further highlighted the noticeable impact of limited deployable hardware on research. As a result, P2 voiced that they were not able to run a memorability study.

> "We did not run a memorability study [for our authentication scheme] mainly due to hardware issues [...] the magical formula would be having an infrastructure that allows to [build hardware-based prototypes] in a very quick way." - **P2**

In summary, experts reported that research involving hardware prototypes often introduces additional challenges. Besides the hardware challenges, many of the challenges voiced by the experts are the result of limited access to appropriate resources and lack of funding, which we discuss further in the context of USEC's research culture in Section 4.5.3.

> **KEY CHALLENGE #2**
>
> Experts voiced that evaluations of USEC systems are expensive and often infeasible to do in an ecologically valid way, especially when they are large-scale and require special equipment or hardware-prototyping experience.

4.3. Sample Size and Selection Process

Sample concerns, especially discussions around the appropriate size of a sample and its characteristics, are highly dominant when conducting experiments and frequently discussed in the HCI (Caine, 2016) and USEC community (Redmiles, Acar, Fahl, & Mazurek, 2017).

4.3.1. Small Sample Sizes

Our interviewees highlighted the importance of collecting large datasets, especially for security evaluations. P3, for example, described the problem with the pool of real-world passwords which is significantly larger compared to a small subset of passwords collected from user studies: *"there's 70 million from a cracked database, you got six and a half thousand – that's like a drop in the ocean" (**P3**)*. P3 also argued that prototype evaluations with small datasets cannot be used to assess security.

> "[we] have got 12-20 users [...] the security data is of no value and the conclusion is that there is no value inside the small sample size." - **P3**

Across all experts there was a consensus that the sample size and selection is a fundamental and ongoing challenge that goes beyond USEC. P11 repeatedly emphasized the challenge with achieving large sample sizes: *"finding a large sample size is really hard" (**P11**)*. While small sample sizes are problematic as pointed out by P3, external factors (e.g., access to different research environments) can have a notable impact on the sampling process and the resulting sample size. P11, for example, voiced that they face significant issues when recruiting participants and that their resources are limited.

> "I recently moved to another country and I was really happy to get 25 [participants] [...] I was really happy to get them but well ..." - **P11**

Throughout the interview, P11 further voiced that for some researchers such a sample is too small and immediately invalidates the conducted research, but that they often overlook the still valuable research and its contributions. P1 voiced that the lack of participants is one of their main research problems and that it is often challenging to convince potential participants to come to the lab.

4.3.2. Biased Recruitment

Additionally to the sample size concerns above, there were discussions and concerns about the recruitment process – the way in which researchers recruit participants for their studies.

> "[sample size and selection] is one of the largest outstanding problems with all HCI systems work which is that we evaluate [our systems] by knocking on the doors of friends and colleagues and be like 'hey, come do my user study and I'll give you $10'." **- P11**

P7 echoed the problem of evaluations using experimenters' social circles and that it is often unclear what happens if the system is evaluated with a more diverse sample and even with people who do not know what privacy is.

> "We run [studies] within our social circles, what happens if we get someone who's elderly, who's not familiar with technology, [or] who doesn't even know what privacy is?" **- P7**

P2 further highlighted that although they have access to a gigantic user pool, which is not comparable to the (often) limited user pools in academic environments, their user pool still runs out and that they still rely on vendors to have access to an even larger set of participants.

> ### KEY CHALLENGE #3
>
> Experts voiced that sample size and recruitment are problems across multiple disciplines and major concerns of USEC research. Small sample sizes and biased participant selection reduce the value and validity of security evaluations.

4.4. Evaluation Methodologies

Discussions on the topic of lab and field studies were common among our participants. We observed themes around the importance of both of them (Section 4.4.1), the value/cost trade-off (Section 4.4.2), and the perceived value of field studies in USEC research (Section 4.4.3).

4.4.1. Importance of Lab and Field Studies

Our experts emphasized the necessity of different evaluation approaches, and that starting with lab studies is often a prerequisite for evaluating USEC systems.

> "There's a place for both [...] I don't think it makes any sense to go directly into the field to evaluate new systems when we haven't done any lab studies at all." **- P9**

There was an agreement across all experts on: 1) lab studies should be conducted before going into the field and 2) the potential of field studies to lead to ecologically valid findings. Experts also voiced that researchers should not underestimate the importance of lab studies. P11, for example, highlighted that different study types come with different pros and cons and that imperfect evaluations of usable privacy and security prototypes can still be valid contributions to the research community.

> "We can have ideas - that's the strength of academia – ideas that are totally radical and new and not going to be evaluated perfectly in the context of a lab study [...] but that doesn't mean they don't have value [or] can't inspire the direction of the usable security and privacy future." **- P11**

P7 further emphasized the importance of taking prototypes out of the lab and putting them into real environments. Other experts mentioned that real-world investigations have a particular importance as participants manifest "demand characteristics'; they subconsciously change their behavior to fit the experimenter's purpose. P11, for example, highlighted the uncertainty about the effect of lab studies on results and that they *"cannot be sure whether [participants] are acting as [they] would act in the wild or if they've changed their behavior because they know they are being part of a study"* **(P11)**.

4.4.2. Value/cost Trade-off

The trade-off between effort in applying a methodology (e.g., lab vs field study) and the value/ecological validity of the corresponding findings was highly discussed by our experts and is considered to be a domain challenge in USEC research (Garfinkel & Lipford, 2014). Our experts stated that lab studies are considered simpler than field studies and that this is one of the main reasons why we see a plethora of lab studies but significantly less field studies in USEC prototyping research.

> "[the] uncharitable view would be that [running lab studies rather than field studies] is just easier to do." **- P8**

> "My take is that it's a mix of convenience, not knowing better, and impossibility as in certain [situations] you can't do [experiments] that are difficult to do that it's not worth the additional effort." **- P2**

P9 also voiced that *"it's easier to do a lab study; the odds of something going wrong are way too high"* **(P9)**. Another expert, P3, voiced that before conducting field studies it is important to compare the value versus the effort.

> "There is a place for [field studies] but is there enough added value in field studies generally that this is important?" **- P3**

Experts also voiced that field studies can be powerful and it is not unlikely that results deviate from lab findings, but researchers need to be clear about what they are seeking rather than being exploratory. P3 elaborated that *"field studies can be valuable but there needs to be a clear value [...] the data will differ from a lab"* **(P3)**. Other experts, for example P4, highlighted the strength of field studies as they provide insights about how systems are really going to be used and how people are going to accept them. On the other hand, P5 voiced it is hard to pinpoint causes of effects in field studies and that achieving accurate results through field studies only is challenging.

KEY CHALLENGE #4

Experts voiced that the choice of evaluation methodology is highly context-dependent and it is important to have a clear vision and expectation of the scale of the evaluation. There is a clear value of field studies; but there is a need of preceding lab studies as pinpointing sources of problems in field studies is otherwise challenging.

4.4.3. Experts' Views on Field Studies in USEC Research

Some experts reported believing that *"field studies are sort of a gold standard"* **(P11)** and that they *"would like to see more about how security fits into real life as opposed*

*to specific little corner cases that are easy to run" (**P9**).* P10 highlighted that the suitability of field studies heavily depends on the required investigation and the legal and ethical considerations, and that this differs a lot between different countries. Experts also described some unsuccessful attempts to study prototypes in a real-world setting.

> "We looked at investigating [our security system] within a real setting but there were just too many legal and ethical constraints around that." **- P2**

P6 and P7 added that such field studies are expensive and that they often have to rely on findings from lab studies only due to budget issues and technological issues they would face in field studies.

<div style="border: 2px solid orange; border-radius: 8px;">

KEY CHALLENGE #5

Experts voiced that legal, ethical, and budget constraints play a major role in decisions around whether to conduct field studies in USEC.

</div>

4.5. USEC's Research Culture

Different research fields and individual researchers come with different sets of behaviors, values, expectations, attitudes, and norms, forming a unique research environment and culture. Open science and reproducibility, for example, are recognized as vital features of science across research fields and considered as a disciplinary norm and value (McNutt, 2014); however, in practice there are significant differences across research communities. Wacharamanotham, Eisenring, Haroz, and Echtler (2020) showed that the process of sharing artifacts is an uncommon practice in the HCI community and Cockburn, Gutwin, and Dix (2018) showed that preregistration has received little to no published attention in HCI whereas other research fields (e.g., psychology) started to award badges for different categories (e.g., "open data", "preregistration") (Eich, 2014), with promising adoption rates in the first year of operation (Nosek et al., 2015).

When it comes to USEC and researchers' behaviors, values, expectations, attitudes, and norms, experts mentioned challenges around the expected, often hard to reach, high ecological validity of usable privacy and security prototype evaluations (Section 4.5.1), USEC researchers' reserved enthusiastic about novel evaluated systems (Section 4.5.2), and the lack of access to research resources (Section 4.5.3). While some of these challenges can also be found in neighboring research communities (e.g., the lack of access to research resources in HCI (Wacharamanotham et al., 2020)), the combination of the challenges and USEC researchers' opinions and their research approaches form a unique research culture.

4.5.1. Towards (high) Ecological Validity

An important objective in usable privacy and security research is to achieve high ecological validity; the extent to which a study adequately reflects real-world conditions. A password study by Fahl, Harbach, Acar, and Smith (2013) showed that participants in a lab study behaved differently compared to their real-world behavior. Although Redmiles et al. (2018) argued that many insights from self-report security data can, when used with care, translate to the real world, they also emphasized that self-reported data can indeed vary from data collected in the field and that alternative research methodologies should be considered for studying detailed constructs. Some of our experts mentioned that USEC researchers often expect high ecological validity

and generalizability of study findings. As a result, they aim to, for example, role-play real-world situations in the lab (Fahl et al., 2013), conduct field studies (e.g., (Harbach, De Luca, & Egelman, 2016; Malkin, Harbach, De Luca, & Egelman, 2017; Mare, Baker, & Gummeson, 2016)), or leverage online studies to increase sample size and target a more representative sample (e.g., (Cheon, Shin, Huh, Kim, & Oakley, 2020; Harbach, De Luca, Malkin, & Egelman, 2016; Markert, Bailey, Golla, Dürmuth, & Aviv, 2020)). However, P12, for example, stated that a real-world evaluation of all systems' usability and security aspects is almost impossible: *"the difficulty in evaluating system security is that the lack of security can have many different sources."* *(P12)*. P12 further emphasized the complexity of security evaluations.

> "All secure systems are alike. But there are many different ways for a system to be not secure. It is not possible to enumerate them all." **- P12**

A concern by P1 was about the lack of common evaluation approaches and that their set of evaluation metrics (e.g., interaction time with a security system, error rate when providing input) often has to evolve from literature reviews because of the lack of any standards. Researchers' various evaluation approaches exacerbate the problem of determining which metrics to investigate and which evaluation method to employ for evaluating USEC systems. P2, for example, voiced that the variety of evaluation approaches often also leads to a wide range of different system evaluations and conclusions.

> "If you look at five different usable security papers you can't compare them because they have used slightly different approaches of evaluating the different parts of their systems [...] you can't really say which one was better or worse." **- P2**

P2 particularly highlighted the subjectivity of privacy and security and that many researchers have strong opinions when it comes to the evaluation. P2 also voiced that the lack of standardized sets to evaluate security schemes makes it even harder to fully address ecological validity and perform comparisons between multiple works.

> "I am not aware of any standard scenarios that can say 'okay, here now we can compare it if we're running a lab study'." **- P2**

In line with Key Challenge #4, P11 underlined the need for a clear vision of what is expected of evaluations that are either conducted in lab settings and are likely less ecologically valid, or are conducted as organized field studies that are still limited to an extent due to research participation effects (Nichols & Maner, 2008; Orne, 1962).

> "We [as a community] just need to be a little bit more open to what sort of solutions/evaluations we are expecting out of [something] that has not actually been deployed in the real world." **- P11**

KEY CHALLENGE #6

Experts voiced that aiming for evaluations with high ecological validity is crucial in USEC research; however, they also mentioned that USEC prototype evaluations are often incapable of achieving high ecological validity.

4.5.2. Creating Space for Novel Solutions

P11 expressed that while problem-scoping research, e.g., identifying usability issues in existing systems, is important, it is equally important to conduct problem-solving

research. Some experts also raised the concern that the USEC community is very focused on the evaluation part when a large element of the contribution is building a system functional enough to demonstrate effectiveness in terms of both deployment and usability.

> "The [USEC community] wants to evaluate everything when like a big part of your contribution is just the fact that you could build this [system]." **- P11**

P11 further argued that without recognition of the value of functional solutions, which may come with limitations imposed by the real world, the community may struggle to really engage with the realities of solving problems.

> "I feel like we as a community refuse to accept that kind of contribution – then you know, we're shooting ourselves in the foot, we're never going to be part of the broader conversation." **- P11**

Other experts discussed the community's focus on realistic use cases resulting in limited enthusiasm for building speculative future-oriented solutions. P8 mentioned that they have seen some shift recently, but problem-scoping and problem-solving USEC research are still not balanced.

> "I like some of the shift we've seen recently [...] to actually really look at finding ways of supporting [users]." **- P8**

Considering the implementation of novel solutions, P11 argued that there is still a lack of future-oriented USEC research where use cases are more speculative or avant-garde.

> "In general the usable security and privacy security community is not very imaginative [...] they don't really like thinking too far in the future." **- P11**

P10 agreed to some extent and voiced that the USEC community does not appreciate research where they have to imagine worlds that do not exist.

KEY CHALLENGE #7

Experts voiced that problem-solving research is relatively scarce in recent USEC research. While problem-scoping lays the foundation for further investigations, research that implements and evaluates usable privacy and security prototype solutions is also valuable as voiced by the experts.

4.5.3. Accessibility and Availability of Resources

Experts highlighted the lack of open-source material within the USEC community that negatively affects their research outcome. According to P6 there is a significant lack of open-source implementations of usable privacy and security systems available. P4 voiced that the lack of open-source material makes it time-consuming and challenging to build certain features and P11 even suggested to collectively build a platform that supports researchers in their research.

> "How can we create a platform that will make it super easy for other researchers to build upon the foundation that you've created?" **- P11**

Experts also voiced that their research is often driven by the hardware that is available. For example, P2 faced challenges in investigating a security system's usability while users are walking.

> "We had the idea of putting people on a treadmill for the evaluation [...] but then didn't

have a treadmill." **- P2**

Building upon the sample size discussions in Section 4.3, experts also asserted that finding a broad user base is even more critical in academia and that this is where most academic studies suffer because the resources for recruiting are limited.

> ### KEY CHALLENGE #8
>
> Experts voiced that the current USEC research community does not consistently support sharing research resources, for example, access to hardware prototypes, software implementations, and platforms for conducting studies.

4.6. Academia & Industry in USEC Research

Getting access to real systems used by companies is challenging and the lack of access can result in lower ecological validity as well as barriers to transitioning research results into practice. For example, one issue in privacy and security research is that potential industry partners are concerned about harmful findings and do not allow any "vulnerability research" (Gamero-Garrido, Savage, Levchenko, & Snoeren, 2017), including prototype-building work. P2 related such an incident.

> "We did have some connections with [companies] but they are like: ´you can't touch our machines'." **- P2**

Experts voiced that this type of research can be of great value, but there are concerns over legal challenges. Building upon the discussions around USEC's research culture, we present experts' comments on the lack of collaborations between academia and industry (Section 4.6.1) and the resulting limited real-world impact (Section 4.6.2).

4.6.1. Academia and Industry - Status Quo

Experts voiced that although there are collaborations between academia and industry, there is still room for improvements when it comes to exchanging knowledge, sharing research resources, and accelerating impact. Our experts voiced that one of the resulting problems is the lack of hardware accessibility (similar to Key Challenge #8) that leads to limited research contributions and therefore decreases ecological validity: *"if they just lent us [an ATM] for a period of time it would have been really good to do our studies" (P8)*. Another expert, P7, also brought up the ATM example and the corresponding challenges with financial institutions.

> "Which bank would allow [to install] some random prototypical hardware; probably no bank." **- P7**

The lack of access to real-world hardware is only one problem according to P8. P8 also voiced that another considerable problem is companies' fear of security leaks that impacts usable privacy and security research.

> "If you are working in the security in an environment where there is real-world security, they often won't let you do any observations and I think that's really bad [...] they are afraid that you're going to find something that means the security isn't working." **- P8**

Besides that, researchers are restricted in publishing findings based on observations within companies: *"I've been lucky to have done observational studies a couple of times [but] I wasn't allowed to publish them" (P8)*. The importance of ecological validity in

USEC underpins the need for more real-world studies where users actually use those systems on a daily basis. One way to conduct more of these investigations is, according to P11, to collaborate closely with industry and establish stronger collaborations.

4.6.2. USEC Research and its "Real-world" Impact

When asking our experts whether or not they see controlled lab studies as the "way to go" to evaluate security and privacy-protecting systems and what progress they would like to see within the USEC community, discussions around the impact of USEC research on real-world applications came up and that this transition, moving USEC research and corresponding usable privacy and security solutions into practice, is still lacking. Some experts voiced that the problem is not that the USEC community lacks ideas for usable and secure systems; instead, they would like to see how these systems and solutions fit into real life. P9, for example voiced that many publications end in a heap of privacy and security schemes that never find their way into users' daily lives.

> "There's a lot of proposed authentication schemes out there and a lot of them aren't gonna move forward like a lot of them are ideas, they didn't really work out, they're not really showing any promise and so you know, we discarded them." **- P9**

Although there are technologies that are widely deployed nowadays (e.g., anti-phishing technology, two-factor authentication) much of USEC research has indeed not been applied in the real world; examples include enhancing authentication on mobile devices (Bianchi, Oakley, Kostakos, & Kwon, 2010; De Luca et al., 2014; Khamis, Hassib, von Zezschwitz, Bulling, & Alt, 2017; von Zezschwitz, De Luca, Brunkow, & Hussmann, 2015) or protecting users' privacy when interacting with public displays (De Luca, von Zezschwitz, Pichler, & Hussmann, 2013; Ragozin et al., 2019; von Zezschwitz et al., 2015). Our experts voiced that a major reason for the limited impact is the huge gap between prototype evaluations and being able to use these systems in real-world settings: *"there's a huge gap between possibility and building the system and commercialization" (**P5**)*. Complementing this, P8 emphasized that many researchers do not want to change their existing theories or their skillset; therefore, there seems to exist some kind of resistance to change within the USEC community. Our experts also highlighted that the interests of USEC researchers and practitioners vary widely. Particularly, some experts were concerned about some other USEC experts mindset.

> "I've seen this in rebuttals [...] when I write a review about something [...] and they are like oh well so many other people have published lab studies, why should I have to go out and do something differently [...] it's a lot harder, it's a lot more work and as long as I can get this stuff published why should I bother?" **- P8**

Experts also highlighted that USEC research should go beyond publications and not be entirely driven by the "publish or perish" mindset (McGrail, Rickard, & Jones, 2006). P11 encouraged the USEC community to think big and collaboratively aim for more than "little projects".

> "How can we make that little project the next like D3.js[1] for usable security?" **- P11**

P12 further criticized the opinionated mindset of many researchers and that the academic career is often considered to be more important than having real-world impact.

---

[1] Bostock, Ogievetsky, and Heer (2011) presented data-driven documents (D3) as a novel approach for visualizations at IEEE Transactions on Visualization and Computer Graphics in 2011. Originating from research conducted at Stanford University, D3.js found its way into web development and is nowadays a library for data-driven visualizations and used by many developers.

"Most researchers' goal is to produce papers and get their degree or tenure; few researchers are [actually] building and deploying working systems." **- P12**

> **KEY CHALLENGE #9**
>
> Experts voiced there is a lack of strong collaborations between academia and industry and that there seems to exist some kind of resistance to changes within the USEC community; resulting in limited real-world impact.

5. Discussion

We have identified 9 key challenges, each of which contributes to answering **RQ₁**: current bottlenecks of USEC research that involves prototyping and user studies are manifold and it is challenging to pinpoint a single source (Key Challenge #1 – #9). In **RQ₂**, we asked what the USEC community needs to better transition research contributions to the real world. We discuss how our findings contribute to **RQ₂**'s answer and provide a discussion of and comparison to similar challenges in neighboring HCI disciplines. To conclude, we discuss the implications of our work and provide ways forward for both individual researchers and the broader USEC community.

5.1. There is no one Best Way for doing USEC Research

Our experts noted that it is impossible to enumerate all security aspects of a system but that imperfectly prototyped and evaluated systems can still have value and inspire the direction of USEC's future. Arguably, the opinions brought up by our experts around the design, development, and evaluation of prototype systems are not far away from the HCI literature. Greenberg and Buxton (2008) and Shneiderman et al. (2016) emphasized that the choice of evaluation methodology should evolve from the actual problem (e.g., what are users' needs) and appropriate research questions. In the context of usable security, it is also important to note here that a system's usability and security oftentimes highly depends on the specific context (e.g., external factors can impact a system's state or a user's behavior (Kainda, Fléchais, & Roscoe, 2010)). The value, benefits, and drawbacks of different evaluation methods were echoed by our interviewees together with the non-trivial part of threat modeling (Key Challenge #1 and #4). Below, we discuss our experts' voiced comments in more detail and tie those back to the broader research field.

5.1.1. Adjusting Expectations of Prototype Developments and Evaluations

According to some of the experts (e.g., see P12's statement in Section 4.5.1 or P11's statements in Section 4.5.1 & 4.5.2), the problem is exacerbated by some researchers' expectation of an exhaustive evaluation that assesses every single aspect of a system's characteristics in an ecologically valid setting. There are many reasons that make this often infeasible when evaluating novel systems, including: 1) the need to run lab studies first to evaluate the new elements in the prototype and pinpoint causes of problems and 2) not having the resources (e.g., hardware) to produce multiple prototypes for in-the-wild testing. The voiced hardware prototyping and ecological validity challenges (Key Challenge #2 and #6) voiced by our experts can also be found in neighbor-

ing research communities such as Ubicomp. Prototyping novel ubiquitous systems is challenging (Dix, Finlay, Abowd, & Beale, 2003; Greenberg & Fitchett, 2001) and often requires additional expertise and specific tools (e.g., knowledge about different electronic components, access to soldering irons). Greenberg and Fitchett (2001) even described developing and combining physical devices and interfacing them within the application software as one of the biggest obstacles. In a similar vein to the lack of sharing research resources and expertise in building hardware voiced by our experts (Key Challenge #2 and #8), Greenberg and Fitchett (2001) observed that researchers who develop systems based on physical devices are often required to start from scratch and face many difficulties. In their own little project, building a reactive media space environment, one of their colleagues (an electrical engineer) joined the team and provided significant support in the hardware-building process (Greenberg & Fitchett, 2001). More than ten years later, we can indeed see similar interdisciplinary collaborations in USEC research. One example is the Back-of-Device prototype by De Luca, von Zezschwitz, Nguyen, et al. (2013) and their follow up work XSide (De Luca et al., 2014). The form factor of their first prototype significantly reduced the generalisability of the results of one-handed interaction. While Greenberg and Fitchett (2001) benefited greatly from an electrical engineer that joined the project, the prototype by De Luca et al. (2014) benefited greatly from the 3D printing expertise of one of the researchers; thus improved, together with an advanced algorithm, user experience.

This shows that **collaborations can greatly improve USEC prototype systems and corresponding evaluations**. As emphasized by Fléchais and Faily (2010), usable privacy and security research requires a variety of researchers from different research areas beyond USEC (e.g., psychology, economics), which we discuss further in Section 5.4.2. The voiced prototype-related challenges (e.g., Key Challenge #2 and #6) also suggest that **expectations of prototype developments and evaluations should be adjusted in situations where building "perfect" prototypes and conducting highly realistic evaluations are too challenging**.

5.1.2. Bridging the Gap between Lab and Field Studies

The USEC community has been debating the respective value of lab and field studies for some time, with our experts similarly mentioning the need to be open to alternative evaluation approaches (Key Challenge #2 and #6). Discussions around lab and field studies, especially when and how field studies are *"worth the hassle"* are also discussed in neighboring communities such as Mobile HCI (Kjeldskov & Skov, 2014). A corresponding critical evaluation and comparison of a lab and field study even impacted the Mobile HCI research field in the subsequent years (Kjeldskov & Skov, 2014; Kjeldskov, Skov, Als, & Høegh, 2004). Kjeldskov et al. (2004) discovered more usability issues in the lab than in a similar field study, for roughly half the cost; consequently, the researchers concluded that the added value of field studies is very little and neglectable, which resulted in a heated debate about the generalisability as the study did not cover long-term use and adoption (Iachello & Terrenghi, 2005).

In USEC research, the long-term use and evaluation of systems is indeed an important component. For example, previous works showed that habituation can impact users' perception and security behavior (e.g., see the replication study of CMU's SSL study (Sotirakopoulos, Hawkey, & Beznosov, 2011; Sunshine, Egelman, Almuhimedi, Atri, & Cranor, 2009)). Other works also emphasized the importance of habituation and its key role in USEC research (e.g., in classification of genuine login attempts (Syed, Banerjee, Cheng, & Cukic, 2011) or in research on security alert dialogs (Mau-

rer, De Luca, & Kempe, 2011)). That being said, Greenberg and Buxton (2008), for example, assert that there is a need to recognize many other appropriate ways to evaluate and validate work and that usability evaluations can be ineffective if naively done "by rule" rather than "by thought" and that "a combination of methods – from empirical to non-empirical to reflective – will likely help triangulate and enrich the discussion of a system's validity." (Greenberg & Buxton, 2008).

In USEC, there seems to be a need to fundamentally rethink current study paradigms (Alt & von Zezschwitz, 2019) and frameworks for understanding privacy risks and solutions in personalization systems (Toch, Wang, & Cranor, 2012). For example, the uptake of smart speakers that could collect sensitive data about users (e.g., (Alrawais, Alhothaily, Hu, & Cheng, 2017; Lau, Zimmerman, & Schaub, 2018; Toch et al., 2012)) requires a change in the way current security and privacy prototype systems are designed and evaluated.

There have been suggestions to improve ecological validity of usability and security evaluations in the lab. For example, role-playing real-world situations (Fahl et al., 2013) to mimic scenarios where security is a secondary task (which is usually the case in the real world (A. M. Sasse et al., 2001)). However, it has also been argued that these approaches can not necessarily compete with the ecological validity of real-world studies and should therefore not be treated as an alternative. As voiced by our experts, the context and expectation of the corresponding evaluation method is important and USEC research needs all facets of evaluation methods, including studies of different types beyond traditional lab and field studies. While preceding lab studies and follow up field studies are vital to transition usable privacy and security prototype systems into practice in the long run, alternative evaluation methods are equally important and can inspire usable privacy and security research in the future.

A potential direction to address the challenges around ecological validity is to leverage novel technologies for prototype development, deployment, and evaluation. As brought up by one of our experts (see also Section 5.4.3), 3D printing can significantly facilitate prototyping of security systems and USEC research in general (as evidenced by, for example, De Luca et al. (2014); Marky et al. (2020)). Future work could also consider the use of online platforms to facilitate field studies. For example, Redmiles et al. (2018) showed that many insights from online surveys on security and privacy translate to the real world. In line with Redmiles et al. (2018), Mazurek et al. (2013) suggest that passwords collected through online studies can be a reasonable proxy for real-world passwords. Similar to the transition of lab to online studies, there has also been a movement in human-centered research to use virtual and augmented reality to conduct user-centered evaluations of, for example, authentication schemes, IoT devices, and public displays (Mäkelä et al., 2020; Mathis, Vaniea, & Khamis, 2021; Voit, Mayer, Schwind, & Henze, 2019). Mathis, Vaniea, and Khamis (2021) showed that virtual reality (VR) can serve as a suitable test bed for the usability and security evaluation of real-world authentication schemes. In a similar vein, Mäkelä et al. (2020) reported that user behavior is largely similar in field studies of public displays compared to behavior in immersive virtual reality, while Voit et al. (2019) compared conducting empirical studies online, in virtual reality, in augmented reality, in the lab, and in in-situ studies to find that some findings are comparable across them while others are not. Following P11's emphasis on aiming for something beyond little projects, building an online platform that is capable of evaluating physical privacy and security systems in an ecologically valid way could be a powerful approach to establish an infrastructure for USEC research that may be complementary to lab and field studies.

One key message here is that **we as a research community should be mindful**

of the challenges that USEC researchers encounter when evaluating usable privacy and security prototype systems. **There is often great value and depth in findings from lab studies.** It is also important to note that **novel technologies and evaluation methods can augment USEC research in the long run** (e.g., the previously discussed 3D printing examples by De Luca et al. (2014) and Marky et al. (2020) or using VR as a test-bed (Mathis, Vaniea, & Khamis, 2021)). It is without question that field studies are essential for high ecological validity; however, sometimes they are infeasible due to constraints beyond researchers' capabilities due to lack of resources or the nature of the prototype (e.g., evaluating a tethered hardware prototype). In these cases, field studies can take place only if the prototype features much higher fidelity than what can be achieved in typical research environments.

5.2. Selecting Sample Sizes in the Presence of Constraints

A major discussion point in our interviews was about sample sizes and selection processes (Key Challenge #3), which is a major domain challenge in USEC research (Garfinkel & Lipford, 2014; Redmiles et al., 2017). Looking at the content of the neighboring HCI community, we see a wide range of sample sizes and compositions used. For example, Caine highlighted that twelve participants was the most common sample size across papers published in CHI 2014 (Caine, 2016). Focusing on usability only, Turner, Lewis, and Nielsen (2006) found that five users allow discovering 80% of a system's usability problems. Similarly to our experts' concerns about the way participant recruitment happens in their USEC research (e.g., *"we evaluate [our systems] by knocking on the doors of friends and colleagues and be like 'hey, come do my user study"' (P11)*, Lazar, Feng, and Hochheiser (2017) argued that there are many HCI studies that come with a small and non-diverse sample (e.g., students only); therefore, often do not allow generalising results. We discuss the "correct" sample size selection and its reality further in Section 5.2.1 and 5.2.2.

5.2.1. "Correct" Sample Size Selection

Classically the "correct" way to decide on a sample size is highly dependent on the research question and the type of data being collected (Lazar et al., 2017; Redmiles et al., 2017). For example, qualitative studies that focus on ground-up approaches use the concept of "saturation" (Guest et al., 2006), where data is collected till the uncovered insights start saturating, i.e., increasing the sample size does not reveal additional insights. Saturation is an interesting approach because sample size is decided while the research is ongoing rather than up-front, making it challenging to know at the start how many participants will be needed. Quantitative studies that involve statistical testing use a very different approach. The number of needed participants is calculated up-front using information like expected variance to perform a power analysis computation of how many subjects are required to reach statistical significance (Field & Hole, 2002; Yatani, 2016). However, this often clashes with the realities of finding and conducting experiments with users.

Redmiles et al. (2017) emphasized the importance of different sampling methods in different research contexts and the need to rely on some form of convenience sampling due to, for example, time and cost considerations. Sample composition is also an issue since some groups, like security engineers, penetration testers, and chief security officers, are not necessarily easy to get time with. Yet, targeting populations like this even at low sample numbers may be the most appropriate approach to answer

a specific research question (Redmiles et al., 2017). The tension harks back to the initial attempts by Nielsen (1994) to find valid ways of conducting usability tests in low-budget environments, such as universities. Approaches like Think Aloud (Nielsen, 1994) and Delphi (Loo, 2002) studies were specifically designed to extract the maximum amount of usability data from small samples. USEC research has some unique properties that make the application of these approaches challenging for prototype testing, namely that security is often a secondary task (Cranor & Garfinkel, 2005; Garfinkel & Lipford, 2014) where users' main goal is likely something other than the prototype's security function. Since many of the "discount" usability approaches focus on having the user engage with the tested system as a primary task, they are challenging to fully adapt to USEC (Kainda et al., 2010).

5.2.2. "Realities" of Sample Size Selections

Our experts voiced that for human-centered privacy and security evaluations and corresponding sample sizes and selections there exist many different opinions within the USEC community. P3, for example, argued that a sample size of 12-20 users for security evaluations is too small to have any value. There are indeed published works that come with noticeable large sample sizes. For example, Ur et al. (2017) conducted an online study with N=4509 participants to detail the security and usability impact of a password meter's design dimensions, Cheon et al. (2020) assessed and evaluated a security framework in large crowd-sourced online studies (N=2619 and N=4000), and Markert et al. (2020) conducted an online study to analyze the security of smartphone unlock PINs with N=1220 participants.

At the same time, security evaluations of published work at top USEC venues such as ACM CHI (Das, Laput, Harrison, & Hong, 2017; De Luca et al., 2014; De Luca, von Zezschwitz, Nguyen, et al., 2013; Khamis et al., 2016; Kim et al., 2010; von Zezschwitz et al., 2015) and USENIX SOUPS (De Luca et al., 2009; Krombholz et al., 2016; Tari, Ozok, & Holden, 2006) studied noticeable smaller samples. On top of that, some security evaluations are even based on a single expert attacker (Bianchi, Oakley, & Kwon, 2011a; De Luca et al., 2009; De Luca, von Zezschwitz, Nguyen, et al., 2013; Krombholz et al., 2016) or on a small sample of trained participants who were put in the role of attackers (Abdrabou, Khamis, Eisa, Ismail, & Elmougy, 2019; Bianchi, Oakley, & Kwon, 2011b; Khamis et al., 2018). The previously mentioned works show the wide range of acceptable participant numbers within USEC and how much that acceptance varies across sub-domains, resulting in no single rule about how many participants and what type of participants (e.g., experts, novices) are needed. Taking shoulder surfing as an example, the spectrum of study designs results in a wide range of types of findings, impacts on the validity, and limits the ability of researchers to compare results and systems (Bošnjak & Brumen, 2020; Wiese & Roth, 2015).

The message here is that **working collectively towards a research standard or a set of roughly defined guidelines could be beneficial for both individual researchers and the USEC research community as a whole**. This could help the USEC community to a) **support early career researchers** in their usable privacy and security research decisions (e.g., which sampling method to apply? how many participants? against which threat should the system protect users?) and b) **facilitate comparisons** between works.

5.2.3. The Quest to Find (many) Participants

As shown in, for example, the USEC works by Aviv, Wolf, and Kuber (2018); Cheon et al. (2020); Felt, Reeder, Almuhimedi, and Consolvo (2014); Harbach, Von Zezschwitz, Fichtner, De Luca, and Smith (2014); Markert et al. (2020), an approach that facilitates achieving large sample sizes is to use crowd-sourcing online platforms such as Amazon Mechanical Turk or establishing university-industry collaborations that allow the investigation of security systems at larger scales. Yet, the deployment and corresponding evaluation of usable privacy and security prototypes still remains a challenge. Online services are often not suitable to, for example, evaluate hardware-based prototypes or prototypes for platforms that participants do not own (e.g., phones with a touch-sensitive rear (De Luca, von Zezschwitz, Nguyen, et al., 2013), smart glasses (Winkler et al., 2015), VR headsets (Mathis, Williamson, Vaniea, & Khamis, 2021)).

In line with the suggestions in Section 5.1.2, our key message and a future research direction here is to **investigate alternative platforms** for conducting research that can **balance** a) reaching out to a **large number of participants**, and b) delivering **realistic experiences to ensure high ecological validity**. A further direction to address this as a community is to facilitate and encourage collaborations across researchers. For example, one of the major concerns voiced by our experts is the lack of resources or research infrastructure to allow them to reach out to many participants, such as limited funding to compensate participants, conduct online studies, or purchase needed hardware (Key Challenge #2). Sharing research resources across research groups, such as prototype systems, evaluation equipment, procedures, and evaluation platforms would benefit the USEC community as a whole. Other fields make their resources available for collaborators. For example, chemistry and physics labs share their research equipment with other groups, and arrange research visits to allow their collaborators to leverage their unique equipment (Wolfgang Glänzel, 2001). This is often done in return for sharing results, intellectual property rights or co-authorship of research outputs which is a win-win for everyone involved.

Finding a suitable sample is crucial for evaluations that should take place before taking USEC concepts to the real world, and thus the suggestions above contribute to answering R_{Q2} (*"What does the USEC community need to better facilitate the transition of artifact contributions into practice?"*). We discuss the potential impact of encouraging collaborations in USEC further in Section 5.4.2.

5.3. Problem-scoping and Problem-solving

To further answer RQ_2, we refer to Figure 1 to make results more tangible. Note that the figure is based on experts' statements and our interpretation of the conducted interviews (Key Challenges #1 - #9). To this end, we distinguish between: ❶ the real world, ❷ problem-scoping research, and ❸ problem-solving research. USEC research puts a strong emphasis on problem-scoping research (e.g., Balebako, Jung, Lu, Cranor, and Nguyen (2013); De Luca et al. (2010); Harbach, De Luca, and Egelman (2016); Harbach et al. (2014); Inglesant and Sasse (2010); Leon et al. (2013); Mare et al. (2016); Markert et al. (2020); Nguyen, Derr, Backes, and Bugiel (2019); Redmiles (2019)) where, for example, users' behavior is observed to identify privacy and security issues. The generated knowledge is then used to inform, teach, and protect people (Althobaiti, Vaniea, & Zheng, 2018; Canova, Volkamer, Bergmann, & Borza, 2014; Kirlappos & Sasse, 2012). However, some of our experts voiced there is relatively less progress in leveraging these findings to also develop novel privacy-protecting and secu-
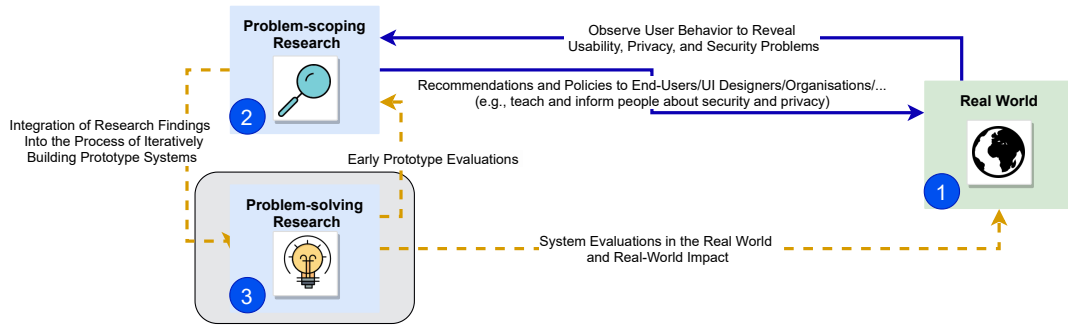
Figure 1. The schematic figure represents a substantial part of conducted work in USEC with a focus on usable privacy and security prototype systems. Dotted orange lines indicate underdeveloped links and solid blue lines strong links.

rity solutions and, more importantly, facilitate their transition into practice. Balancing both research directions has the potential to result in noticeable real-world impact in the long run. P11 emphasized the importance of investing in problem-solving because otherwise, as they put it: *"we're shooting ourselves in the foot and never going to be part of the broader conversation"*. The USEC community clearly values how systems are used, not just how they are built, so while problem-solving needs more emphasis (Key Challenge #7), it should not be at the expense of proper usability, privacy, and security evaluations. Conducting human-centered studies should be an integral part of the evaluation of usable privacy and security systems rather than just "box ticking", and should be integrated at an early stage of the process. That being said, while some communities appreciate innovative solutions even if they lack in-depth user evaluations, a novel privacy or security system that is not usable will not be secure; if a system is not usable, users will work around it or misuse it, resulting in poor privacy and security (Adams & Sasse, 1999; Whitten & Tygar, 1999). Besides users' perceived ease of use and social benefits that can influence users' adoption decision, the potential risks associated to other users' privacy also play an important role in the likelihood of use and adoption of ubiquitous systems (Rauschnabel et al., 2016).

The message here is that **it is important to collectively understand what the people who are finally going to use the systems need before building many different prototype systems that end in publications but do not contribute to the bigger picture: transition research into practice and provide users with usable privacy and security systems**. While spotting users' privacy and security issues is essential, it is equally important to integrate these findings into an iterative research process and build solutions that eventually find their way into practice and solve some of those issues. This would also foster collaborations between researchers who conduct problem-scoping research and those who conduct problem-solving research, and would eventually close the loop depicted in Figure 1.

5.4. Open Research and Collaborations to Mitigate Challenges

There is a number of ways in which collaborations can help address some of the named challenges.

5.4.1. Accessibility of Research Material

The segregation of research material (e.g., prototypes) from publications is common and makes reproducibility and comparisons challenging (Sugrim, Liu, McLean, & Lindqvist, 2019; Vines et al., 2014; Wicherts, Borsboom, Kats, & Molenaar, 2006). While individual researchers of the HCI and security communities are gradually implementing elements of the Open Science movement (Innovation, 2016), there are still many contributions that are often not publicly available. A recent CHI paper by Wacharamanotham et al. (2020) shows that sharing artifacts is uncommon in HCI research, with percentages between 14% for raw selective data (e.g., notes during ethnographic studies) and 47% for hardware (e.g., 3D designs, circuit diagrams). This lack of accessibility makes it challenging to build on previous works or compare findings. The USEC community can encourage this by setting guidelines that encourage open source (e.g., sharing 3D models or circuit diagrams of prototypes), including their accessibility as a criteria for acceptance, or make additions to the reviewing process through, for example, the inclusion of badges (Kay, Haroz, Guha, Dragicevic, & Wacharamanotham, 2017; Kidwell et al., 2016). This would help mitigate Key Challenge #8. However, it is equally important to take a look at the reasons for not sharing research material. There are situations where researchers face restrictions that are beyond their capabilities. For example, as reported by Wacharamanotham et al. (2020), the top two reasons for not sharing research data are (1) the sensitivity of data and (2) the lack of permission. Wacharamanotham et al. (2020) even found that sharing research artifacts may sometimes even be prohibited by researchers' respective institutional review board (IRB) or ethics board. Restrictions introduced by institutional regulations or industry partners should not, at any point, disadvantage individual researchers and restrict them from publishing their research. While Open Science is important, it is also important to note that properly preparing all elements of the research for public consumption requires a lot of time, and is less rewarded in academia compared to conducting research and publishing papers (McGrail et al., 2006). Whether or not the act of "publishing papers" should be researchers' core task is another question, one that our experts have only partially touched (see Section 4.6.2).

5.4.2. Collaborations across Research Groups

Our experts emphasized that interdisciplinary research could contribute to addressing the lack of resources and the faced hardware challenges when developing novel USEC prototypes (Key Challenge #2 and #8). P8 highlighted the need and value of *"collaborations between usable security people and the people who are close to building [systems] and can create different variants" (P8)*. It has to be said that there are successful collaborations across research groups that resulted in fruitful privacy and security research, with the privacy icon research by Cranor and Schaub (2020), now used by California law (Tkacik, 2020), as one of the most recent examples that involved researchers from different universities with different backgrounds including privacy, software, and law research. In a similar vein, when it comes to usable privacy and security prototypes and their evaluation, one way to mitigate the challenges of reaching out to participants could be by establishing strong collaborations among research groups. For example, if a consortium of research groups collectively builds an infrastructure that facilitates participant recruitment, it would help the involved researchers and the USEC community as a whole. Looking at more distributed models of participant recruitment, including potential access to target-specific, hard-to-reach user groups, and establishing an infrastructure that allows sharing research equipment

could also help mitigate Key Challenge #8.

5.4.3. Engagement with Industry and Transition to Practice

Our experts voiced that building hardware prototypes is challenging and is often out of their expertise. P8, for example, highlighted that the USEC community needs more collaborations with prototype-building experts. Sometimes the industry is better equipped to build prototypes, or has resources (e.g., possibility of reaching millions of users (Felt et al., 2014)) that researchers in academia do not posses. Among the successful examples in the USEC community is Felt et al. (2014) who collaborated with Google to collect data based on 130,754 user interactions, which would have been challenging using academic resources only. However, it is important to note here that university-industry collaborations can be complex and introduce further challenges (e.g., bureaucracy or the inflexibility of universities (Schofield, 2013)) that may impact the success of such collaborations (Rybnicek & Königsgruber, 2019). The lack of transition into practice (Key Challenge #9) is according to our experts also attributable to the mindset of many researchers – "publish or perish" (McGrail et al., 2006), which leads to many prototype systems that are sufficient for user studies and publications, but not necessarily deployable in a real-world setting. While the transition of USEC prototype systems into practice has been endorsed by our experts, it is also important to revisit the fundamental idea of prototyping in human-centered research. While providing users with both usable and secure systems is one of the primary goals of USEC research, and therefore, it seems to be important to transition USEC prototype systems into practice, USEC research is much more. Focusing solely on the transition into practice would greatly undermine USEC's goals. In a broader sense, as put by Garfinkel and Lipford (2014):

> "The goal of academic research in usable security should be to help speed the discovery (and therefore the adoption) of techniques that simultaneously improve both usability and security." - (Garfinkel & Lipford, 2014, p.4)

In USEC research, it is often both: transitioning usable and secure systems into practice but also using the research around the prototype for educational purposes and to facilitate learning. The prototype system by De Luca, von Zezschwitz, Nguyen, et al. (2013), for example, can be interpreted as a novel authentication method with the aim to become a product, or it can be construed as a system built to evaluate how usable and secure authentication can be if we integrate the back of a device for user authentication.

There are also prototype systems that were not built to transition into practice, but rather to shed light on users' perception of different authentication concepts on doors (Mecke, Pfeuffer, Prange, & Alt, 2018) or on mobile devices (Prange, Mecke, Nguyen, Khamis, & Alt, 2020). USEC researchers have also built prototype systems to investigate the extent to which security systems from mobile devices can be adapted for virtual reality applications (George et al., 2017) or to study the impact of different input techniques and threat models on users' security (Mathis, Williamson, et al., 2021). In these cases, the primary research goal is not necessarily to transition the prototype into practice but rather to make significant contributions to USEC's research field, facilitate learning, and inspire potential future research. In fact, experts also voiced that industry involvement and collaborations are required to transform early usable privacy and security systems into actual deployable systems. The lack of transition into practice is indeed also a key challenge in Ubicomp research (Caceres &

Friday, 2012; Davies & Gellersen, 2002). Shneiderman (2016) emphasized the importance of collaborations and that we all should combine practical problems with the development of theory because each supports and drives the other.

In a similar vein, USEC research involves both *problem-scoping* and *problem-solving* USEC research that is often treated independent of each other and seems to be not balanced (Section 5.3 and Key Challenge #7). As also mentioned by Fléchais and Faily (2010), one way the USEC community could foster collaborations between academia and industry is by introducing industry tracks to conferences, and creating forums that bring USEC researchers together with potential industry partners (e.g., DS3[3]). To further spark interest in usable privacy and security research and emphasize its relevance, additional conference-independent online events (e.g., tutorials, seminars) similar to, for example, the Quarterly Workshop on Security Information Workers[4] could be organized. This would contribute towards mitigating Key Challenge #5 and #9. Engagement with industry could also help to overcome some of experts' voiced legal and ethical constraints (e.g., having access to specific study settings). Ethical and legal considerations are a fundamental part of usable privacy and security research (e.g., see the security field study of ATM use by De Luca et al. (2010)). That being said, Ethics also forms one of the seven HCI grand challenges (Stephanidis et al., 2019) and is a vital component of modern research in general (Yip, Han, & Sng, 2016).

6. Concluding Remarks

Although some of our key challenges are more relevant to USEC (e.g., threat modeling, USEC's research culture), many issues USEC experts face when designing, prototyping, and evaluating usable privacy and security systems can also be found in neighboring research communities such as HCI, Mobile HCI, Ubicomp. Is this surprising? No, not at all. While the birth of usable privacy and security happened around 1995 with the works by Zurko and Simon (1996), Whitten and Tygar (1999), Adams and Sasse (1999), and Jermyn et al. (1999), the first formal gathering of the USEC community can indeed be traced back to a workshop at a non security-focused venue: ACM CHI 2003 (Andrew, Chris, & Flinn, 2003). The history of usable privacy and security research, including the way in which the USEC community has been established and our experts' voiced challenges, shows that USEC research does not exist in a vacuum. In fact, usable privacy and security has borrowed many research methods from the HCI community and neighboring communities (Garfinkel & Lipford, 2014). Johnston, Eloff, and Labuschagne (2003), for example, used and refined the ten usability heuristics by Nielsen (2005) to promote and enable security awareness of users when interacting with computer systems. HCI as a discipline has established many more widely-used guidelines, toolkits, and processes to incorporate usability into products at an early stage such as the seven stages of action by Norman (1988) or the eight golden rules of interface design by Shneiderman and Plaisant (2010). The inherently interdisciplinary nature of usable privacy and security and the challenges around security evaluations, threat modeling, and ecological validity, and the lack of strong links between problem-scoping and problem-solving USEC research (see Figure 1) make usable privacy and

---

[3]The Developing Secure Systems Summit (DS3, https://ds3summit.github.io/, accessed 06/03/2021) seeks to establish a new meeting ground for researchers and practitioners from software industry, academia, research labs, and governments.

[4]The WSIW workshop (https://wsiw.sec.uni-hannover.de/, accessed 06/03/2021) is a quarterly event and aims to develop and stimulate discussion about security information worker.

security research unique, hard, and often impedes the transition of USEC systems into practice. As put by Fléchais and Faily (2010):

> "progress in usable security research and design has been slow, due in part to the need to master a large amount of (usually) mutually exclusive, yet necessary, skills and knowledge." - (Fléchais & Faily, 2010, p.1)

By synthesizing opinions from USEC experts that have not seen in-depth discussions in prior literature, and raising awareness of the challenges when prototyping and evaluating usable privacy and security systems, we hope to provide a common starting point for ongoing discussions within the USEC community.

Lessons Learned and Ways Forward. To summarize the lessons learned from our work, we outline five ways forward that can be tackled on the individual researcher level together with community efforts to strengthen the links between problem-scoping research, problem-solving research, and the real world, highlighted in Figure 1:

1) consider different evaluation methods and be realistic about what conclusions can be made from each paradigm;
2) establish new evaluation paradigms to cope with the challenges outlined above with representative samples;
3) consider how we, on a researcher and USEC community level, can establish procedures and structures that strengthen collaborations across academics and between academic research labs and industry;
4) share research resources (e.g., making prototypes open source) and enable other researchers to access research data (e.g., raw data from studies) to increase the research impact in the long run; and
5) balance a) *problem-scoping* and b) *problem-solving* USEC research with proper user-centered evaluations.

7. Conclusion

A substantial part of usable privacy and security is to iteratively design, implement, and evaluate prototype systems that address usability, security, and privacy. However, providing users with prototypes that are usable while also fulfilling their privacy or security objectives is still a major challenge. In this work, we reported on twelve semi-structured interviews with established and nascent researchers from academia and industry who have published research that evaluates usable privacy and security prototypes in human-centered studies. We synthesized their opinions of challenges encountered when conducting this type of research, discussed the challenges in the light of neighboring communities, and identified five ways forward researchers and the USEC community as a whole can pursue to mitigate these challenges.

ciety of Edinburgh (award number #65040).

About the Authors

Florian Mathis is a Ph.D. candidate at the University of Glasgow and the University of Edinburgh. His research is in human-computer interaction, usable privacy and security, and virtual reality. He is interested in exploring how virtual reality can better support the development and evaluation of usable privacy and security systems.

Dr Kami Vaniea received her Ph.D. in computer science at Carnegie Mellon University and is now a Lecturer/Assistant Professor in Cyber Security at the University of Edinburgh. Her research is in human factors of cyber security and privacy aiming to better understand the protection needs of all types of users.

Dr Mohamed Khamis received his Ph.D. in media informatics at LMU Munich and is now a Lecturer/Assistant Professor at the University of Glasgow. His research is at the intersection of ubiquitous computing and privacy. He is interested in understanding ubiquitous technologies' privacy implications and in developing novel privacy-protecting systems.

8. Appendices

Appendix A. Interview Invitation

We asked USEC experts if they are willing to be interviewed about their research and included links to several of their research papers that we identified in the review described in Section 3.1.
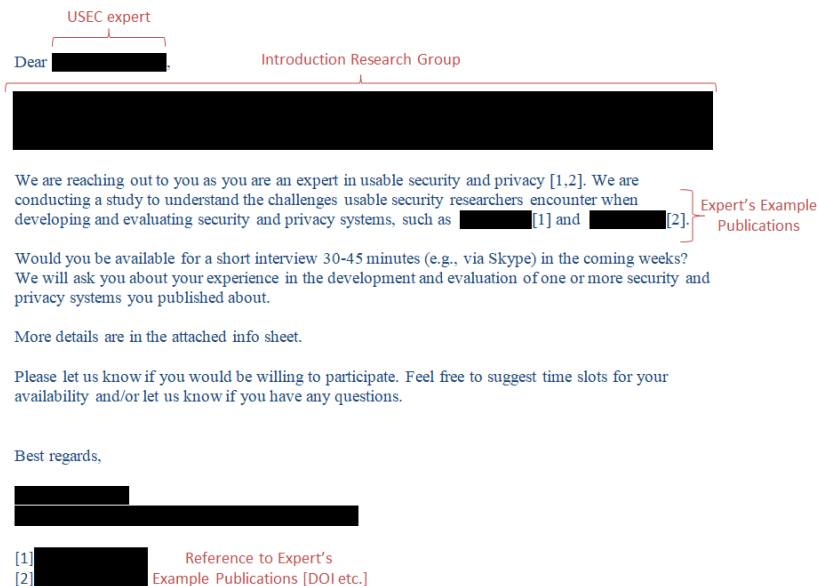

Figure A1. Our interview request included an introduction of our research group, example papers of the expert, and an attached information sheet. Note that we censored specific parts in the email for anonymity reasons.

Appendix B. Semi-structured Interview Questions

(1) **Typical Research Journey from Idea to Publication**
  (a) Let us consider a novel security or privacy-preserving system: If we walk along the path, from an initial idea to the final publication, how would these steps look like?
  (b) With a focus on each specific step: What are challenges or limitations that you encountered when designing, implementing, and evaluating such prototype systems?

(2) **Research Challenges and Limitations**
  (a) Were there limitations that you encountered when iteratively designing, implementing, and evaluating prototype systems?
  (b) What were the most challenging parts when developing *[experts' prototype system]*? Were there any limitations or things you would have preferred to do differently but could not do so?
  (c) What are your thoughts regarding the approaches USEC researchers apply to evaluate privacy and security?

(3) **The Ecological Validity of Current Evaluations**
  (a) Do you see *controlled lab studies* as the "way to go" to evaluate security and privacy-aware prototype systems?
  (b) What are your thoughts on the different study types (e.g., lab, online, or in-the-wild studies) USEC researchers currently apply to assess a prototype system's privacy/security and usability?
  (c) What keeps USEC researchers and practitioners away from investigating security and privacy-aware systems in more realistic contexts (e.g., at a public space such as a bus station)?
  (d) Would you prefer to see *"more realistic"* studies, for example, field studies? Can you please outline why or why not you think so?
  (e) Talking about the ecological validity of human-centered evaluations: What conditions have in your opinion a significant influence on the validity of research findings?
  (f) Let's assume you have the time and resources available to re-run parts of your *[papers study]* again. Would there be anything you would like to investigate in addition to the metrics you have already mentioned in your publications?

References

Abdrabou, Y., Khamis, M., Eisa, R. M., Ismail, S., & Elmougy, A. (2019). Just gaze and wave: Exploring the use of gaze and gestures for shoulder-surfing resilient authentication. In *Proceedings of the 11th acm symposium on eye tracking research & applications.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3314111.3319837

Acar, Y., Fahl, S., & Mazurek, M. L. (2016, Nov). You are not your developer, either: A research agenda for usable security and privacy research beyond end users. In *2016 ieee cybersecurity development* (p. 3-8). Retrieved from https://doi.org/10.1109/SecDev.2016.013

Adams, A., & Sasse, M. A. (1999, December). Users Are Not the Enemy. *Communications of the ACM*, *42*(12), 40–46. Retrieved from https://doi.org/10.1145/322796.322806

Alrawais, A., Alhothaily, A., Hu, C., & Cheng, X. (2017, Mar). Fog computing for the internet of things: Security and privacy issues. *IEEE Internet Computing*, *21*(2), 34-42. Retrieved from https://doi.org/10.1109/MIC.2017.37

Alt, F., & von Zezschwitz, E. (2019). Emerging trends in usable security and privacy. *Journal of Interactive Media*, *18*(3), 189–195.

Althobaiti, K., Vaniea, K., & Zheng, S. (2018, April). Faheem: Explaining urls to people using a slack bot. In *Symposium on digital behaviour intervention for cyber security.*

Andrew, P., Chris, L., & Flinn, S. (2003, April). Workshop on human-computer interaction and security systems. In *Chi 2003.* Fort Lauderdale, Florida, USA: Association for Computing Machinery. Retrieved from http://www.andrewpatrick.ca/CHI2003/HCISEC/

Aviv, A. J., Wolf, F., & Kuber, R. (2018). Comparing video based shoulder surfing with live simulation. In *Proceedings of the 34th annual computer security applications conference* (p. 453–466). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3274694.3274702

Balebako, R., Jung, J., Lu, W., Cranor, L. F., & Nguyen, C. (2013). "little brothers watching you": Raising awareness of data leaks on smartphones. In *Proceedings of the ninth symposium on usable privacy and security.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2501604.2501616

Baudisch, P., Sinclair, M., & Wilson, A. (2006). Soap: A pointing device that works in mid-air. In *Proceedings of the 19th annual acm symposium on user interface software and technology* (p. 43–46). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/1166253.1166261

Bianchi, A., & Oakley, I. (2016). Wearable authentication: Trends and opportunities. *it-Information Technology*, *58*(5), 255–262.

Bianchi, A., Oakley, I., Kostakos, V., & Kwon, D. S. (2010). The Phone Lock: Audio and Haptic Shoulder-Surfing Resistant PIN Entry Methods for Mobile Devices. In *Proceedings of the fifth international conference on tangible, embedded, and embodied interaction* (p. 197–200). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/1935701.1935740

Bianchi, A., Oakley, I., & Kwon, D. S. (2011a). Spinlock: A single-cue haptic and audio pin input technique for authentication. In E. W. Cooper, V. V. Kryssanov, H. Ogawa, & S. Brewster (Eds.), *Haptic and audio interaction design* (pp. 81–90). Berlin, Heidelberg: Springer Berlin Heidelberg.

Bianchi, A., Oakley, I., & Kwon, D. S. (2011b). Using mobile device screens for authentication. In *Proceedings of the 23rd australian computer-human interaction conference* (p. 50–53). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2071536.2071542

Bošnjak, L., & Brumen, B. (2020). Shoulder surfing experiments: A systematic literature review. *Computers & Security*, 102023.

Bostock, M., Ogievetsky, V., & Heer, J. (2011). D3: Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics*. Retrieved from http://vis.stanford.edu/papers/d3

Brudy, F., Ledo, D., Greenberg, S., & Butz, A. (2014). Is Anyone Looking? Mitigating Shoulder Surfing on Public Displays through Awareness and Protection. In *Proceedings of the international symposium on pervasive displays* (p. 1–6). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2611009.2611028

Caceres, R., & Friday, A. (2012). Ubicomp systems at 20: Progress, opportunities, and challenges. *IEEE Pervasive Computing*, *11*(1), 14-21. Retrieved from https://doi.org/10.1109/MPRV.2011.85

Caine, K. (2016). Local Standards for Sample Size at CHI. In *Proceedings of the 2016 chi conference on human factors in computing systems* (p. 981–992). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2858036.2858498

Canova, G., Volkamer, M., Bergmann, C., & Borza, R. (2014). NoPhish: an anti-phishing

education app. In *International workshop on security and trust management* (p. 188–192).

Chen, Y., Li, H., Teng, S.-Y., Nagels, S., Li, Z., Lopes, P., . . . Zheng, H. (2020). Wearable microphone jamming. In *Proceedings of the 2020 chi conference on human factors in computing systems* (p. 1–12). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3313831.3376304

Cheon, E., Shin, Y., Huh, J., Kim, H., & Oakley, I. (2020, may). Gesture authentication for smartphones: Evaluation of gesture password selection policies. In *2020 ieee symposium on security and privacy (sp)* (p. 327-345). Los Alamitos, CA, USA: IEEE Computer Society. Retrieved from https://doi.ieeecomputersociety.org/10.1109/SP.2020.00034

Chun, S. Y., Kang, J., Kim, H., Lee, C., Oakley, I., & Kim, S. (2016). Ecg based user authentication for wearable devices using short time fourier transform. In *2016 39th international conference on telecommunications and signal processing (tsp)* (p. 656-659).

Cockburn, A., Gutwin, C., & Dix, A. (2018). Hark no more: On the preregistration of chi experiments. In *Proceedings of the 2018 chi conference on human factors in computing systems* (p. 1–12). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3173574.3173715

Corbin, J., & Strauss, A. (2014). *Basics of qualitative research: Techniques and procedures for developing grounded theory.* Sage publications.

Cranor, L. F., & Garfinkel, S. (2005). *Security and usability: designing secure systems that people can use.* " O'Reilly Media, Inc.".

Cranor, L. F., & Schaub, F. (2020, October). How to (in)effectively convey privacy choices with icons and link text. In *2020 USENIX conference on privacy engineering practice and respect (PEPR 20).* USENIX Association. Retrieved from https://www.usenix.org/conference/pepr20/presentation/cranor

Das, S., Laput, G., Harrison, C., & Hong, J. I. (2017). Thumprint: Socially-Inclusive Local Group Authentication Through Shared Secret Knocks. In *Proceedings of the 2017 chi conference on human factors in computing systems* (p. 3764–3774). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3025453.3025991

Davies, N., & Gellersen, H. . (2002). Beyond prototypes: challenges in deploying ubiquitous systems. *IEEE Pervasive Computing*, *1*(1), 26-35. Retrieved from https://doi.org/10.1109/MPRV.2002.993142

De Luca, A., Denzel, M., & Hussmann, H. (2009). Look into my eyes! can you guess my password? In *Proceedings of the 5th symposium on usable privacy and security.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/1572532.1572542

De Luca, A., Harbach, M., von Zezschwitz, E., Maurer, M.-E., Slawik, B. E., Hussmann, H., & Smith, M. (2014). Now You See Me, Now You Don't: Protecting Smartphone Authentication from Shoulder Surfers. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 2937–2946). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2556288.2557097

De Luca, A., Langheinrich, M., & Hussmann, H. (2010). Towards understanding atm security: A field study of real world atm use. In *Proceedings of the sixth symposium on usable privacy and security.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/1837110.1837131

De Luca, A., von Zezschwitz, E., Nguyen, N. D. H., Maurer, M.-E., Rubegni, E., Scipioni, M. P., & Langheinrich, M. (2013). Back-of-device authentication on smartphones. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 2389–2398). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2470654.2481330

De Luca, A., von Zezschwitz, E., Pichler, L., & Hussmann, H. (2013). Using Fake Cursors to Secure On-Screen Password Entry. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 2399–2402). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2470654.2481331

Dix, A., Finlay, J., Abowd, G. D., & Beale, R. (2003). *Human-computer interaction.* Pearson Education.

Eich, E. (2014). *Business not as usual.* Sage Publications Sage CA: Los Angeles, CA.

Fahl, S., Harbach, M., Acar, Y., & Smith, M. (2013). On the ecological validity of a password study. In *Proceedings of the ninth symposium on usable privacy and security.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2501604.2501617

Fallman, D. (2003). Design-oriented human-computer interaction. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 225–232). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/642611.642652

Felt, A. P., Reeder, R. W., Almuhimedi, H., & Consolvo, S. (2014). Experimenting at Scale with Google Chrome's SSL Warning. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 2667–2670). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2556288.2557292

Field, A., & Hole, G. (2002). *How to design and report experiments.* Sage.

Fléchais, I., & Faily, S. (2010). Security and usability: Searching for the philosopher's stone.

Gamero-Garrido, A., Savage, S., Levchenko, K., & Snoeren, A. C. (2017). Quantifying the pressure of legal risks on third-party vulnerability research. In *Proceedings of the 2017 acm sigsac conference on computer and communications security* (p. 1501–1513). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3133956.3134047

Garfinkel, S., & Lipford, H. R. (2014). Usable security: History, themes, and challenges. *Synthesis Lectures on Information Security, Privacy, and Trust*, *5*(2), 1–124.

George, C., Khamis, M., Buschek, D., & Hussmann, H. (2019, March). Investigating the third dimension for authentication in immersive virtual reality and in the real world. In *2019 ieee conference on virtual reality and 3d user interfaces (vr)* (p. 277-285). Retrieved from https://doi.org/10.1109/VR.2019.8797862

George, C., Khamis, M., von Zezschwitz, E., Burger, M., Schmidt, H., Alt, F., & Hussmann, H. (2017). Seamless and secure vr: Adapting and evaluating established authentication systems for virtual reality. In *Proceedings of the network and distributed system security symposium (ndss 2017).* NDSS. Retrieved from http://dx.doi.org/10.14722/usec.2017.23028

Gong, L., Lomas, M. A., Needham, R. M., & Saltzer, J. H. (1993, June). Protecting poorly chosen secrets from guessing attacks. *IEEE Journal on Selected Areas in Communications*, *11*(5), 648-656. Retrieved from https://doi.org/10.1109/49.223865

Greenberg, S., & Buxton, B. (2008). Usability evaluation considered harmful (some of the time). In *Proceedings of the sigchi conference on human factors in computing systems* (p. 111–120). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/1357054.1357074

Greenberg, S., & Fitchett, C. (2001). Phidgets: Easy development of physical interfaces through physical widgets. In *Proceedings of the 14th annual acm symposium on user interface software and technology* (p. 209–218). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/502348.502388

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? an experiment with data saturation and variability. *Field methods*, *18*(1), 59–82.

Harbach, M., De Luca, A., & Egelman, S. (2016). The anatomy of smartphone unlocking: A field study of android lock screens. In *Proceedings of the 2016 chi conference on human factors in computing systems* (p. 4806–4817). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2858036.2858267

Harbach, M., De Luca, A., Malkin, N., & Egelman, S. (2016). Keep on lockin' in the free world: A multi-national comparison of smartphone locking. In *Proceedings of the 2016 chi conference on human factors in computing systems* (p. 4823–4827). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2858036.2858273

Harbach, M., Von Zezschwitz, E., Fichtner, A., De Luca, A., & Smith, M. (2014). It's a hard lock life: A field study of smartphone (un)locking behavior and risk perception. In *Proceedings of the tenth usenix conference on usable privacy and security* (p. 213–230). USA: USENIX Association.

Hayashi, E., Riva, O., Strauss, K., Brush, A. J. B., & Schechter, S. (2012). Goldilocks and the two mobile devices: Going beyond all-or-nothing access to a device's applications. In *Proceedings of the eighth symposium on usable privacy and security.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2335356.2335359

Hoda, R., Noble, J., & Marshall, S. (2011). Grounded theory for geeks. In *Proceedings of the 18th conference on pattern languages of programs.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2578903.2579162

Houde, S., & Hill, C. (1997). Chapter 16 - what do prototypes prototype? In M. G. Helander, T. K. Landauer, & P. V. Prabhu (Eds.), *Handbook of human-computer interaction (second edition)* (Second Edition ed., p. 367-381). Amsterdam: North-Holland. Retrieved from https://www.sciencedirect.com/science/article/pii/B9780444818621500820

Hundlani, K., Chiasson, S., & Hamid, L. (2017). No passwords needed: The iterative design of a parent-child authentication mechanism. In *Proceedings of the 19th international conference on human-computer interaction with mobile devices and services.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3098279.3098550

Iachello, G., & Hong, J. (2007). End-user privacy in human–computer interaction. *Foundations and Trends in Human–Computer Interaction*, *1*(1), 1-137. Retrieved from http://dx.doi.org/10.1561/1100000004

Iachello, G., & Terrenghi, L. (2005). Mobile hci 2004: Experience and reflection. *IEEE Pervasive Computing*, *4*(1), 88-91.

Inglesant, P. G., & Sasse, M. A. (2010). The true cost of unusable password policies: Password use in the wild. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 383–392). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/1753326.1753384

Innovation, O. (2016). Open science, open to the world–a vision for europe. *European Comission*.

Ishii, H., & Ullmer, B. (1998, 09). Tangible bits: Towards seamless interfaces between people, bits and atoms. *Conference on Human Factors in Computing Systems - Proceedings*. Retrieved from https://doi.org/10.1145/258549.258715

Jermyn, I., Mayer, A., Monrose, F., Reiter, M. K., & Rubin, A. D. (1999). The Design and Analysis of Graphical Passwords. In *Proceedings of the 8th conference on usenix security symposium - volume 8* (p. 1). USA: USENIX Association.

Johnston, J., Eloff, J. H. P., & Labuschagne, L. (2003, December). Features: Security and human computer interfaces. *Comput. Secur.*, *22*(8), 675–684. Retrieved from https://doi.org/10.1016/S0167-4048(03)00006-3

Kainda, R., Fléchais, I., & Roscoe, A. W. (2010). Security and usability: Analysis and evaluation. In *2010 international conference on availability, reliability and security* (p. 275-282).

Kawakita, J. (1991). The original kj method. *Tokyo: Kawakita Research Institute*, *5*.

Kay, M., Haroz, S., Guha, S., Dragicevic, P., & Wacharamanotham, C. (2017). Moving transparent statistics forward at chi. In *Proceedings of the 2017 chi conference extended abstracts on human factors in computing systems* (p. 534–541). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3027063.3027084

Khamis, M., Alt, F., Hassib, M., von Zezschwitz, E., Hasholzner, R., & Bulling, A. (2016). Gazetouchpass: Multimodal authentication using gaze and touch on mobile devices. In *Proceedings of the 34th annual acm conference extended abstracts on human factors in computing systems.* New York, NY, USA: ACM.

Khamis, M., Hassib, M., von Zezschwitz, E., Bulling, A., & Alt, F. (2017). Gazetouchpin: Protecting sensitive data on mobile devices using secure multimodal authentication. In

Proceedings of the 19th acm international conference on multimodal interaction. New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/3136755.3136809

Khamis, M., Trotter, L., Mäkelä, V., Zezschwitz, E. v., Le, J., Bulling, A., & Alt, F. (2018, December). Cueauth: Comparing touch, mid-air gestures, and gaze for cue-based authentication on situated displays. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, *2*(4). Retrieved from https://doi.org/10.1145/3287052

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L.-S., ... others (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLoS biology*, *14*(5), e1002456. Retrieved from https://doi.org/10.1371/journal.pbio.1002456

Kim, D., Dunphy, P., Briggs, P., Hook, J., Nicholson, J. W., Nicholson, J., & Olivier, P. (2010). Multi-touch authentication on tabletops. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 1093–1102). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/1753326.1753489

Kirlappos, I., & Sasse, A. M. (2012, March). Security education against phishing: A modest proposal for a major rethink. *IEEE Security Privacy*, *10*(2), 24-32. Retrieved from https://doi.org/10.1109/MSP.2011.179

Kjeldskov, J., & Skov, M. B. (2014). Was it worth the hassle? ten years of mobile hci research discussions on lab and field evaluations. In *Proceedings of the 16th international conference on human-computer interaction with mobile devices &; services* (p. 43–52). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2628363.2628398

Kjeldskov, J., Skov, M. B., Als, B. S., & Høegh, R. T. (2004). Is it worth the hassle? exploring the added value of evaluating the usability of context-aware mobile systems in the field. In S. Brewster & M. Dunlop (Eds.), *Mobile human-computer interaction - mobilehci 2004* (pp. 61–73). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from https://doi.org/10.1007/978-3-540-28637-0_6

Krombholz, K., Hobel, H., Huber, M., & Weippl, E. (2015). Advanced social engineering attacks. *Journal of Information Security and Applications*, *22*, 113 - 122. Retrieved from http://www.sciencedirect.com/science/article/pii/S2214212614001343 (Special Issue on Security of Information and Networks)

Krombholz, K., Hupperich, T., & Holz, T. (2016, June). Use the force: Evaluating force-sensitive authentication for mobile devices. In *Twelfth symposium on usable privacy and security (SOUPS 2016)* (pp. 207–219). Denver, CO: USENIX Association. Retrieved from https://www.usenix.org/conference/soups2016/technical-sessions/presentation/krombholz

Lau, J., Zimmerman, B., & Schaub, F. (2018, November). Alexa, Are You Listening? Privacy Perceptions, Concerns and Privacy-Seeking Behaviors with Smart Speakers. *Proceedings of the ACM on Human-Computer Interaction*. Retrieved from https://doi.org/10.1145/3274371

Lazar, J., Feng, J. H., & Hochheiser, H. (2017). *Research methods in human-computer interaction*. Morgan Kaufmann.

Leon, P. G., Ur, B., Wang, Y., Sleeper, M., Balebako, R., Shay, R., ... Cranor, L. F. (2013). What matters to users? factors that affect users' willingness to share information with online advertisers. In *Proceedings of the ninth symposium on usable privacy and security*. New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2501604.2501611

Liu, R., Cornelius, C., Rawassizadeh, R., Peterson, R., & Kotz, D. (2018, March). Vocal resonance: Using internal body voice for wearable authentication. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, *2*(1). Retrieved from https://doi.org/10.1145/3191751

Loo, R. (2002). The delphi method: a powerful tool for strategic management. *Policing: An International Journal of Police Strategies & Management*. Retrieved from https://doi.org/10.1108/13639510210450677

Lopes, P., You, S., Cheng, L.-P., Marwecki, S., & Baudisch, P. (2017). Providing Haptics to Walls & Heavy Objects in Virtual Reality by Means of Electrical Muscle Stimulation. In *Proceedings of the 2017 chi conference on human factors in computing systems* (p. 1471–1482). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3025453.3025600

Lopes, P., You, S., Ion, A., & Baudisch, P. (2018). Adding Force Feedback to Mixed Reality Experiences and Games Using Electrical Muscle Stimulation. In *Proceedings of the 2018 chi conference on human factors in computing systems.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3173574.3174020

Mäkelä, V., Rivu, S. R. R., Alsherif, S., Khamis, M., Xiao, C., Borchert, L. M., ... Alt, F. (2020). Virtual Field Studies: Conducting Studies on Public Displays in Virtual Reality. In *Proceedings of the 38th annual acm conference on human factors in computing systems.* New York, NY, USA: ACM. Retrieved from http://dx.doi.org/10.1145/3313831.3376796

Malkin, N., Harbach, M., De Luca, A., & Egelman, S. (2017, January). The anatomy of smartphone unlocking: Why and how android users around the world lock their phones. *GetMobile: Mobile Comp. and Comm.*, *20*(3), 42–46. Retrieved from https://doi.org/10.1145/3036699.3036712

Mare, S., Baker, M., & Gummeson, J. (2016, June). A study of authentication in daily life. In *Twelfth symposium on usable privacy and security (SOUPS 2016)* (pp. 189–206). Denver, CO: USENIX Association. Retrieved from https://doi.org/10.5555/3235895.3235912

Markert, P., Bailey, D. V., Golla, M., Dürmuth, M., & Aviv, A. J. (2020, May). This PIN Can Be Easily Guessed: Analyzing the Security of Smartphone Unlock PINs. In *Ieee symposium on security and privacy* (pp. 1525–1542). San Francisco, California, USA: IEEE.

Marky, K., Schmitz, M., Zimmermann, V., Herbers, M., Kunze, K., & Mühlhäuser, M. (2020). 3d-auth: Two-factor authentication with personalized 3d-printed items. In *Proceedings of the 2020 chi conference on human factors in computing systems* (p. 1–12). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3313831.3376189

Mathis, F., Vaniea, K., & Khamis, M. (2021). Replicueauth: Validating the use of a lab-based virtual reality setup for evaluating authentication systems. In *Proceedings of the 2021 chi conference on human factors in computing systems.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3411764.3445478

Mathis, F., Williamson, J., Vaniea, K., & Khamis, M. (2020). Rubikauth: Fast and secure authentication in virtual reality. In *Proceedings of the 38th annual acm conference extended abstracts on human factors in computing systems.* New York, NY, USA: ACM. Retrieved from http://dx.doi.org/10.1145/3334480.3382827

Mathis, F., Williamson, J., Vaniea, K., & Khamis, M. (2021, January). Fast and secure authentication in virtual reality using coordinated 3d manipulation and pointing. *ACM Transactions on Computer-Human Interaction (ToCHI)*, *1*(28). Retrieved from https://doi.org/10.1145/3428121

Maurer, M.-E., De Luca, A., & Kempe, S. (2011). Using data type based security alert dialogs to raise online security awareness. In *Proceedings of the seventh symposium on usable privacy and security.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2078827.2078830

Mazurek, M. L., Komanduri, S., Vidas, T., Bauer, L., Christin, N., Cranor, L. F., ... Ur, B. (2013). Measuring password guessability for an entire university. In *Proceedings of the 2013 acm sigsac conference on computer & communications security* (p. 173–186). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2508859.2516726

McGrail, M. R., Rickard, C. M., & Jones, R. (2006). Publish or perish: A systematic review of interventions to increase academic publication rates. *Higher Education Research & Development*, *25*(1), 19–35.

McNutt, M. (2014). *Reproducibility.* American Association for the Advancement of Science.

Mecke, L., Pfeuffer, K., Prange, S., & Alt, F. (2018). Open sesame! user perception of physical, biometric, and behavioural authentication concepts to open doors. In *Proceedings of the 17th international conference on mobile and ubiquitous multimedia* (p. 153–159). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3282894.3282923

Meho, L. I. (2006). E-mail interviewing in qualitative research: A methodological discussion. *Journal of the American Society for Information Science and Technology*, *57*(10), 1284-1295. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.20416

Mhaidli, A., Venkatesh, M. K., Zou, Y., & Schaub, F. (2020). Listen only when spoken to: Interpersonal communication cues as smart speaker privacy controls. *Proceedings on Privacy Enhancing Technologies*, *2020*(2), 251 - 270. Retrieved from https://content.sciendo.com/view/journals/popets/2020/2/article-p251.xml

Nguyen, D. C., Derr, E., Backes, M., & Bugiel, S. (2019). Short text, large effect: Measuring the impact of user reviews on android app security & privacy. *2019 IEEE Symposium on Security and Privacy (SP)*, 555-569.

Nichols, A. L., & Maner, J. K. (2008). The good-subject effect: Investigating participant demand characteristics. *The Journal of general psychology*, *135*(2), 151–166.

Nielsen, J. (1994). *Usability engineering.* Morgan Kaufmann.

Nielsen, J. (2005). *Ten usability heuristics.*

Norman, D. A. (1988). *The psychology of everyday things.* Basic books.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., ... others (2015). Promoting an open research culture. *Science*, *348*(6242), 1422–1425.

Ogunyemi, A. A., Lamas, D., Lárusdóttir, M. K., & Loizides, F. (2019). A systematic mapping study of hci practice research. *International Journal of Human–Computer Interaction*, *35*(16), 1461-1486. Retrieved from https://doi.org/10.1080/10447318.2018.1541544

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American psychologist*, *17*(11), 776.

Perez, A. J., Zeadally, S., Matos Garcia, L. Y., Mouloud, J. A., & Griffith, S. (2018). Facepet: Enhancing bystanders' facial privacy with smart wearables/internet of things. *Electronics*, *7*(12), 379.

Prange, S., Mecke, L., Nguyen, A., Khamis, M., & Alt, F. (2020). Don't use fingerprint, it's raining! how people use and perceive context-aware selection of mobile authentication. In *Proceedings of the international conference on advanced visual interfaces.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3399715.3399823

Ragozin, K., Pai, Y. S., Augereau, O., Kise, K., Kerdels, J., & Kunze, K. (2019). Private Reader: Using Eye Tracking to Improve Reading Privacy in Public Spaces. In *Proceedings of the 21st international conference on human-computer interaction with mobile devices and services.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3338286.3340129

Rauschnabel, P. A., Hein, D. W., He, J., Ro, Y. K., Rawashdeh, S., & Krulikowski, B. (2016). Fashion or technology? a fashnology perspective on the perception and adoption of augmented reality smart glasses. *i-com*, *15*(2), 179 - 194. Retrieved from https://www.degruyter.com/view/journals/icom/15/2/article-p179.xml

Redmiles, E. M. (2019, may). "should i worry?" a cross-cultural examination of account security incident response. In *2019 2019 ieee symposium on security and privacy (sp)* (p. 1107-1121). Los Alamitos, CA, USA: IEEE Computer Society. Retrieved from https://doi.ieeecomputersociety.org/10.1109/SP.2019.00059

Redmiles, E. M., Acar, Y., Fahl, S., & Mazurek, M. L. (2017). *A summary of survey methodology best practices for security and privacy researchers* (Tech. Rep.).

Redmiles, E. M., Zhu, Z., Kross, S., Kuchhal, D., Dumitras, T., & Mazurek, M. L. (2018).

Asking for a friend: Evaluating response biases in security user studies. In *Proceedings of the 2018 acm sigsac conference on computer and communications security* (p. 1238–1255). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3243734.3243740

Reilly, D., Salimian, M., MacKay, B., Mathiasen, N., Edwards, W. K., & Franz, J. (2014). Secspace: Prototyping usable privacy and security for mixed reality collaborative environments. In *Proceedings of the 2014 acm sigchi symposium on engineering interactive computing systems* (p. 273–282). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2607023.2607039

Remland, M. S., Jones, T. S., & Brinkman, H. (1995). Interpersonal distance, body orientation, and touch: Effects of culture, gender, and age. *The Journal of Social Psychology*, *135*(3), 281-297. Retrieved from https://doi.org/10.1080/00224545.1995.9713958 (PMID: 7650932)

Rybnicek, R., & Königsgruber, R. (2019). What makes industry–university collaboration succeed? a systematic review of the literature. *Journal of Business Economics*, *89*(2), 221–250.

Saldaña, J. (2015). *The coding manual for qualitative researchers*. Sage.

Saltzer, J. H., & Schroeder, M. D. (1975, Sep.). The protection of information in computer systems. *Proceedings of the IEEE*, *63*(9), 1278-1308.

Sasse, A. M., Brostoff, S., & Weirich, D. (2001, Jul 01). Transforming the 'weakest link' — a human/computer interaction approach to usable and effective security. *BT Technology Journal*, *19*(3), 122–131. Retrieved from https://doi.org/10.1023/A:1011902718709

Sasse, M. A., Smith, M., Herley, C., Lipford, H., & Vaniea, K. (2016). Debunking security-usability tradeoff myths. *IEEE Security and Privacy*, *14*(5), 33–39. Retrieved from http://ieeexplore.ieee.org/abstract/document/7676175/

Saunders, B., Kitzinger, J., & Kitzinger, C. (2015). Anonymising interview data: challenges and compromise in practice. *Qualitative Research*, *15*(5), 616-632. Retrieved from https://doi.org/10.1177/1468794114550439 (PMID: 26457066)

Schaub, F., Könings, B., Lang, P., Wiedersheim, B., Winkler, C., & Weber, M. (2014). Prical: Context-adaptive privacy in ambient calendar displays. In *Proceedings of the 2014 acm international joint conference on pervasive and ubiquitous computing* (p. 499–510). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2632048.2632087

Schofield, T. (2013). Critical success factors for knowledge transfer collaborations between university and industry. *Journal of Research Administration*, *44*(2), 38–56.

Scott, C. R. (2005). Anonymity in applied communication research: Tensions between irbs, researchers, and human subjects. *Journal of Applied Communication Research*, *33*(3), 242-257. Retrieved from https://doi.org/10.1080/00909880500149445

Shneiderman, B. (2016). *The new abcs of research: Achieving breakthrough collaborations*. Oxford University Press.

Shneiderman, B., & Plaisant, C. (2010). *Designing the user interface: strategies for effective human-computer interaction*. Pearson Education India.

Shneiderman, B., Plaisant, C., Cohen, M., Jacobs, S., Elmqvist, N., & Diakopoulos, N. (2016). *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (6th ed.). Pearson.

Sotirakopoulos, A., Hawkey, K., & Beznosov, K. (2011). On the challenges in usable security lab studies: Lessons learned from replicating a study on ssl warnings. In *Proceedings of the seventh symposium on usable privacy and security*. New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2078827.2078831

Stephanidis, C. C., Salvendy, G., of the Group Margherita Antona, M., Chen, J. Y. C., Dong, J., Duffy, V. G., . . . Zhou, J. (2019). Seven hci grand challenges. *International Journal of Human–Computer Interaction*, *35*(14), 1229-1269. Retrieved from https://doi.org/10.1080/10447318.2019.1619259

Sugrim, S., Liu, C., McLean, M., & Lindqvist, J. (2019). Robust performance metrics for

authentication systems. In *Network and distributed systems security*. Retrieved from https://doi.org/10.14722/ndss.2019.23351

Sunshine, J., Egelman, S., Almuhimedi, H., Atri, N., & Cranor, L. F. (2009). Crying wolf: An empirical study of ssl warning effectiveness. In *Usenix security symposium* (pp. 399–416).

Syed, Z., Banerjee, S., Cheng, Q., & Cukic, B. (2011). Effects of user habituation in keystroke dynamics on password security policy. In *2011 ieee 13th international symposium on high-assurance systems engineering* (pp. 352–359).

Tao, H., & Adams, C. (2008, Sep). Pass-Go: A Proposal to Improve the Usability of Graphical Passwords. *International Journal of Network Security*, *7*(2), 273-292. Retrieved from https://doi.org/10.6633/IJNS.200809.7(2).18

Tari, F., Ozok, A. A., & Holden, S. H. (2006). A comparison of perceived and real shoulder-surfing risks between alphanumeric and graphical passwords. In *Proceedings of the second symposium on usable privacy and security* (p. 56–66). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/1143120.1143128

Tkacik, D. (2020). *Cylab researchers design privacy icon to be used by california law.* Retrieved 2020-012-17, from https://www.cylab.cmu.edu//news/2020/12/11-donotsell.html

Toch, E., Wang, Y., & Cranor, L. (2012, 04). Personalization and privacy: A survey of privacy risks and remedies in personalization-based systems. *User Modeling and User-Adapted Interaction*, *22*.

Trowbridge, A., Sharevski, F., & Westbrook, J. (2018). Malicious user experience design research for cybersecurity. In *Proceedings of the new security paradigms workshop* (p. 123–130). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3285002.3285010

Turner, C. W., Lewis, J. R., & Nielsen, J. (2006). Determining usability test sample size. *International encyclopedia of ergonomics and human factors*, *3*(2), 3084–3088.

Ur, B., Alfieri, F., Aung, M., Bauer, L., Christin, N., Colnago, J., ... Melicher, W. (2017). Design and evaluation of a data-driven password meter. In *Proceedings of the 2017 chi conference on human factors in computing systems* (p. 3775–3786). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3025453.3026050

van den Hoonaard, W. C. (2003, Jun 01). Is anonymity an artifact in ethnographic research? *Journal of Academic Ethics*, *1*(2), 141–151. Retrieved from https://doi.org/10.1023/B:JAET.0000006919.58804.4c

Vines, T. H., Albert, A. Y., Andrew, R. L., Débarre, F., Bock, D. G., Franklin, M. T., ... Rennison, D. J. (2014). The availability of research data declines rapidly with article age. *Current biology*, *24*(1), 94–97. Retrieved from https://doi.org/10.1016/j.cub.2013.11.014

Voit, A., Mayer, S., Schwind, V., & Henze, N. (2019). Online, VR, AR, Lab, and In-Situ: Comparison of Research Methods to Evaluate Smart Artifacts. In *Proceedings of the 2019 chi conference on human factors in computing systems.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3290605.3300737

Volkamer, M., Gutmann, A., Renaud, K., Gerber, P., & Mayer, P. (2018, August). Replication study: A cross-country field observation study of real world PIN usage at atms and in various electronic payment scenarios. In *Fourteenth symposium on usable privacy and security (SOUPS 2018)* (pp. 1–11). Baltimore, MD: USENIX Association. Retrieved from https://www.usenix.org/conference/soups2018/presentation/volkamer

von Zezschwitz, E., De Luca, A., Brunkow, B., & Hussmann, H. (2015). Swipin: Fast and secure pin-entry on smartphones. In *Proceedings of the 33rd annual acm conference on human factors in computing systems* (p. 1403–1406). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2702123.2702212

Wacharamanotham, C., Eisenring, L., Haroz, S., & Echtler, F. (2020). Transparency of CHI Research Artifacts:Results of a Self-Reported Survey. In *Proceedings of the 38th annual acm conference on human factors in computing systems.* New York, NY, USA: ACM. Retrieved from http://dx.doi.org/10.1145/3313831.3376448

Whitten, A., & Tygar, J. D. (1999). Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0. In *Proceedings of the 8th conference on usenix security symposium - volume 8* (p. 14). USA: USENIX Association.

Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American psychologist*, *61*(7), 726.

Wiedenbeck, S., Waters, J., Birget, J.-C., Brodskiy, A., & Memon, N. (2005). Authentication Using Graphical Passwords: Effects of Tolerance and Image Choice. In *Proceedings of the 2005 symposium on usable privacy and security* (p. 1–12). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/1073001.1073002

Wiese, O., & Roth, V. (2015). Pitfalls of shoulder surfing studies. NDSS Workshop on Usable Security. Retrieved from https://doi.org/10.14722/usec.2015.23007

Winkler, C., Gugenheimer, J., De Luca, A., Haas, G., Speidel, P., Dobbelstein, D., & Rukzio, E. (2015). Glass unlock: Enhancing security of smartphone unlocking through leveraging a private near-eye display. In *Proceedings of the 33rd annual acm conference on human factors in computing systems* (p. 1407–1410). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2702123.2702316

Wobbrock, J. O., & Kientz, J. A. (2016, April). Research contributions in human-computer interaction. *Interactions*, *23*(3), 38–44. Retrieved from https://doi.org/10.1145/2907069

Wolfgang Glänzel. (2001). National characteristics in international scientific co-authorship relations. *Scientometrics*, *51*(1), 69-115. Retrieved from https://doi.org/10.1023/A:1010512628145

Yatani, K. (2016). Effect sizes and power analysis in hci. In J. Robertson & M. Kaptein (Eds.), *Modern statistical methods for hci* (pp. 87–110). Cham: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-319-26633-6_5

Yip, C., Han, N.-L. R., & Sng, B. L. (2016). Legal and ethical issues in research. *Indian Journal of Anaesthesia*, *60*(9), 684.

Zeng, E., & Roesner, F. (2019). Understanding and improving security and privacy in multi-user smart homes: a design exploration and in-home user study. In *28th {USENIX} security symposium ({USENIX} security 19)* (pp. 159–176). Retrieved from https://www.usenix.org/conference/usenixsecurity19/presentation/zeng

Zurko, M. E., & Simon, R. T. (1996). User-Centered Security. In *Proceedings of the 1996 workshop on new security paradigms* (p. 27–33). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/304851.304859