Position Paper: Replication Crisis in Human-Centered Security Research—Are We There Yet?

Juliane Schmüser*, Jan-Ulrich Holtgrave*, Florian Schaub†, Sascha Fahl*
*CISPA Helmholtz Center for Information Security, Germany,
{juliane.schmueser,jan-ulrich.holtgrave,sascha.fahl}@cispa.de
†University of Michigan, USA, fschaub@umich.edu

***Abstract*—Replicability is critical for meta-science, verifying scientific knowledge, and advancing science as a whole. However, the broader security community and the human-centered security (HCS) research community, in particular, struggle with making studies replicable, and only a few replication studies have been published so far. In this position paper, we analyze Call for Papers (CfPs) and existing replication studies to identify pain points that hinder replication in HCS research, use those insights to discuss whether the community is already in a replication crisis, and assess the extent to which this contributes to or hinders meta-science in the field. We conclude with calls to action to improve reporting transparency and data sharing, better recognize replication studies as valuable scientific contributions, and develop community-driven understandings of how replications should be conducted for HCS research.**

1. Introduction

Under the term *Science of Security*, the security research community, including the HCS community, discusses the field's scientific maturity, methods, and results. At its core, scholars ask whether computer security research is indeed a scientific discipline, i.e., whether it generates fundamental principles and theories rigorously tested through well-designed experiments, or whether the field focuses mostly on observation and engineering. This critical discourse began almost two decades ago in an NSF/IARPA/NSA workshop [1]. The workshop was motivated by a perceived lack of progress in advancing the scientific understanding of fundamental principles in computer security, despite significant engineering success, and aimed to establish and foster research on scientific questions in computer security. It was followed by a special issue of the IEEE Security & Privacy magazine on the topic [2] and the introduction of the Systematization of Knowledge (SoK) paper category to the Call for Paper (CfP) of the IEEE Symposium on Security and Privacy in 2010. SoK papers are described as contributions that systematize, evaluate, and potentially challenge existing knowledge, thus providing a high value to the research community [3]. From there, the debate gained traction again in 2017 with the publication of Herley and Oorschot' SoK paper on the science of security. The authors attested that the field as a whole has failed to learn from principles on which other scientific disciplines have long since formed a consensus. In particular, they criticized HCS research for its tendency to accept proposed alternative hypotheses after a single experiment successfully refutes a corresponding null hypothesis without further investigation or replication. The authors attributed the replication crisis in psychology in part to poor null-hypothesis testing practices [4].

This begs the question of whether the HCS research community is close to, or already in, a *replication crisis*. Recent years have seen several institutionalized attempts to improve scientific rigor (e.g., Artifact Evaluation and Badges, or USENIX Security's Open Science appendix). However, the processes for determining whether artifacts are functional and results are reproducible are primarily optimized for software, systems, and theoretical artifacts; but they are poorly aligned with user studies, which are common in HCS research [5]. In addition, replication, as defined by the ACM as the ability of other researchers to reproduce a study's results using their own experimental setup [6], is difficult to incorporate into the double-blind peer-review process. In HCS research, replication studies remain rare—only a few studies labeled as replications have been published over the past 13 years. This leaves unclear which results from 25 years of HCS research are replicable or generalizable beyond the initial one-off studies.

In this paper, we explore where the HCS research community stands regarding replications, what might hinder them, and whether the field is in a replication crisis. We illustrate that the main challenges are a lack of transparent reporting and incentives to conduct replication studies.

2. Transparency Gap

One major obstacle to successful replication is often the lack of transparent methods reporting in research papers describing an original study. Details on the process and artifacts of the study are necessary for both reproducibility (different researchers, same setup [6]) and replicability (different researchers, different setup [6]). For the former, this is obvious, as the details are needed to faithfully recreate the method. However, for the latter, the details are almost equally important, since they are needed to identify factors

that may have influenced the results and can explain differences. Good reporting transparency is therefore an important precondition to reproducibility and replicability.

Two recent studies by Klemmer et al. identified significant deficits in transparent research reporting in HCS research [5], [7]. In their analysis of published HCS papers [7], Klemmer et al. found that critical artifacts that are crucial for replications are often missing in published papers: interview guides for 25.6% of studies, survey instruments for 20.9%, experiment materials for 20.7% and analyzed software for more than half (55.1%).

Other materials that the community seemingly considers less important, but are still critical to enable replication attempts, are omitted at even higher rates. For example, the recruitment procedure strongly influences and may bias the participant sample. It is therefore crucial to either replicate it closely or be able to identify even subtle differences that may influence a study's outcome. However, while 89.8% of studies in Klemmer et al.'s dataset described their recruitment procedure, 87.6% failed to provide the recruitment materials (such as flyers, email texts, etc.) [7], the design of which can be an important factor.

While these numbers are already concerning, they gain further weight in light of the qualitative insights from Klemmer et al.'s other study investigating researchers' practices and experiences with transparency in HCS [5]. This study does not focus on replications but provides interesting and worrying insight into the effect of the numbers described above on the practicality of replications in HCS. Several of the participants reported that a lack of transparency, whether in missing artifacts or in the description of the study procedure, had impeded attempts to replicate an HCS study [5]. In addition to the quantitative literature analysis, which could only evaluate if a material or information was publicly available, the interview participants' accounts show that the quality of shared materials can be an additional problem, and that reaching out to authors to obtain missing information is not always successful [5].

Taking these findings into account, it is obvious that deficits in the transparency of research reporting contribute to the limited number of replications in HCS research.

3. Replication Reality in HCS

To assess replication quality and style in HCS research, we investigated venues that accept HCS replications and what replications have been published there. Furthermore, we summarize HCS-specific challenges when conducting replications, including a lack of incentives and appreciation in the community for replications, insufficient guidance on what to replicate and how, and a prevalence of qualitative methods and findings that are not easily replicable.

3.1. Publication Venues for HCS Replications

Few top-tier venues explicitly encourage and accept HCS replication studies. We systematically analyzed the

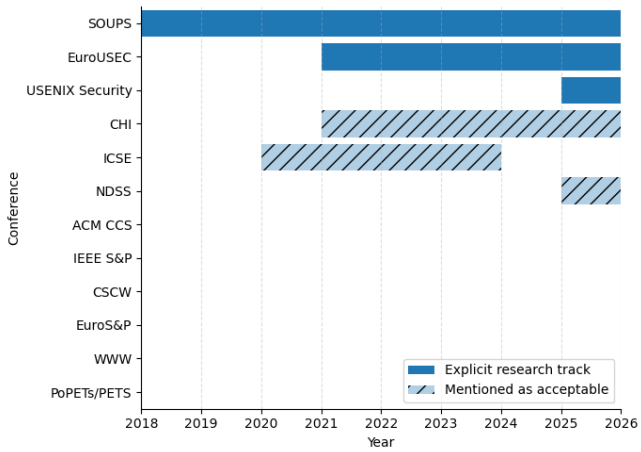


Figure 1. Mentions of replications in CfPs of common HCS venues from 2018–2026.

CfPs for the 12 common HCS venues analyzed by Klemmer et al. [7] from 2018 to 2026 to determine whether replications are encouraged or explicitly acknowledged as an accepted contribution (see Figure 1). Some venues offer *explicit research tracks* for replications in their CfP (e.g., SOUPS 2019 stating: *"confirming previous findings is also considered a contribution; replication focuses on the result while methods could differ and up-to-date methods should be used*"). Others *mention replications* in their CfP or reviewer documentations as an acceptable study type but do not further refer to how they are handled (e.g., CHI 2021 referring to replications in their subcommittee page as: *"The subcommittee [...] welcomes a broad range of contributions, including but not limited to: [...] Empirical study, including replication studies*").

Only SOUPS, USENIX Security, and EuroUSEC have actively been calling for replication papers in at least one of their CfPs since 2018. EuroUSEC has elicited replications in their CfPs since 2021. USENIX Security introduced the special submission type *Replication and Reproduction* in 2025 and continued it in 2026. SOUPS has elicited replications since 2018. However, the CfP remains vague about the type of replications that are welcome or needed, as the CfP only refers to *"important previously published studies*" [8].

CHI, ICSE, and NDSS mention replications on their websites as a study type that can be published at the venue, but they differ significantly in how they address it. CHI does not mention replications in its main CfP, but addresses them in additional documentation about subcommittee selection and links to a document encouraging reviewers to consider replication studies. ICSE included replications in their reviewer guidelines from 2022 to 2024 and also listed *"extensive replication of prior work*" as an acceptable type of contribution. However, for 2025 and 2026, we found no mentions of replications in any of its documents. NDSS is a somewhat special case, as its 2025 and 2026 CfPs mention replications in the context of SoKs, stating that these may

also include *"new experiments to replicate and compare previous solutions.*" Thus, replication studies at NDSS might be accepted as SoKs. Overall, replication studies in HCS have only recently been embraced by more top-ranked venues.

3.2. Published Replications and Their Findings

There are a few published replications of HCS studies, and they report mixed results. We used an extended version of the HCS dataset of Klemmer et al. [7] to identify HCS replication studies by using the title prefix of venues (e.g., *Replication:* for SOUPS) as an identifier and doing a keyword search on the abstracts of each paper.

In total, we identified 14 HCS studies that state they are a replication of a prior study (see Table 1), 12 of which were published at SOUPS and two at EuroUSEC. Nearly all of them deviated from the original study protocol in some way, e.g., by using different recruitment channels, recruiting a different population, or modifying the protocol. Five of the replications confirm the results of the original study, often adding additional insights. Four studies partially replicated the previous results, and five studies obtained results that significantly differed from those of the original study. In one case, this was not unexpected because a sample with greater security expertise was studied [9], but the other three studies used very similar study designs to the original studies. One noteworthy case is the study *"On the challenges in usable security lab studies*" by Sotirakopoulos et al., which confirmed the original study's results in a similar lab experiment, but obtained drastically different results in a field study, casting doubt on the ecological validity of the lab experiment's findings [10].

In addition to these labeled replications, there may be other unlabeled replication studies. Venues such as CHI and ICSE do not require a title prefix for replications, and SOUPS and EuroUSEC do not have a uniform naming convention for these studies (see Table 1). Furthermore, it may not be apparent from the title or abstract whether a paper is a replication of a prior study. Even though we could not systematically identify such studies through our keyword search, we know of some instances from our experience in the field. For example, Atwater et al. [11] replicated the study of Ruoti et al. [12], but neither the title nor the abstract mentions replication, only speaking of a *"study revisit*" in their abstract.

Furthermore, some studies have adopted or used setups from previous studies but leave open whether they consider themselves a replication. For example, Holtgrave et al. [13] used Farke et al.'s experiment setup [14] for the experimental part of their study with a different population and compared their results with the original study but did not describe it as a replication attempt. Similarly, Danilova et al. [15] used a study design adapted from Naiakshina et al. [16], with a slightly modified study population, and published it as a replication study. Yet other studies, such as Zhuo et al.'s [17], shared a lot of the setup with a previous study but are rather explicit about not being a replication due to various design changes (*"[...], our study is neither a replication nor an extension of their work.*") [17].

TABLE 1. AN OVERVIEW OF 14 HCS REPLICATIONS PUBLISHED IN THE PAST TWO DECADES AT SOUPS AND EUROUSEC.

Paper Title	Venue	Ref.
On the challenges in usable security lab studies: lessons learned from replicating a study on SSL warnings	SOUPS'11	[10]
Leading Johnny to water: designing for usability and trust	SOUPS'15	[11]
Replication: Challenges in Using Data Logs to Validate Phishing Detection Ability Metrics	SOUPS'17	[18]
Replication Study: A Cross-Country Field Observation Study of Real World PIN Usage at ATMs and in Various Electronic Payment Scenarios	SOUPS'18	[19]
Replication: No One Can Hack My Mind Revisiting a Study on Expert and Non-Expert Security Practices and Advice	SOUPS'19	[20]
"If you want, I can store the encrypted password": A Password-Storage Field Study with Freelance Developers	SOUPS'20	[15]
Replication: Why We Still Can't Browse in Peace: On the Uniqueness and Reidentifiability of Web Browsing Histories	SOUPS'20	[21]
Replication: Measuring user perceptions in smartphone security and privacy in Germany	EuroUSEC'21	[22]
Replicating a study of ransomware in Germany	EuroUSEC'21	[23]
Replication: Effects of Media on the Mental Models of Technical Users	SOUPS'21	[9]
Replication: How Well Do My Results Generalize Now? The External Validity of Online Privacy and Security Surveys	SOUPS'22	[24]
Replication: Stories as Informal Lessons about Security	SOUPS'22	[25]
Replication: The Effect of Differential Privacy Communication on German Users' Comprehension and Data Sharing Attitudes	SOUPS'22	[26]
Replication: *"No one can hack my mind*" - 10 years later: An update and outlook on experts and non-experts security practices and advice	SOUPS'25	[27]

Our analysis demonstrates that there is currently no consensus in the HCS community for how to conduct, describe, and label replication studies. At the same time, even among the few labeled replication studies we analyzed, multiple of them provide evidence that contradicts the original study's findings, underlining the importance of encouraging and fostering more HCS replication studies. However, not being able to replicate a previous result does not mean that the results from the initial authors are wrong but is merely an indication that other effects than anticipated might be at play or that the investigated landscape has changed since the initial study.

3.3. Challenges for HCS Replications

The fact that only a few venues explicitly encourage replications and the small number of published HCS repli-

cations highlight that conducting replications, which can eventually support meta-science to comprehensively evaluate evidence across studies, is challenging for HCS research. Several key factors inhibit conducting replication studies.

First, although the results from Klemmer et al. show that replications are highly valued by many members of the community, there are few incentives to actually conduct replication studies [5]. Some of Klemmer et al.'s interviewees reported that they struggled to get replication papers published or did not even attempt them because they considered them insufficiently rewarded. They further noted that publication was particularly difficult for studies that confirm previous results and are thus perceived to lack novelty, which many venues and reviewers appear to value highly [5]. This tendency may introduce a severe publication bias.

Second, the lack of research transparency discussed in Section 2 can make it very difficult to conduct replication studies and meaningfully compare results with an original study.

Third, it is unclear which studies are valuable to replicate for the community. Even venues that encourage replications are vague about what kind of replications would constitute a sufficient contribution for publication. Venues with replication tracks, like SOUPS, refer to replications of *"important previously published studies*", leaving open which kinds of studies are considered *"important*" [8].

Finally, the lack of explicit encouragement of replication studies at many venues may make it advantageous for authors to *not* have their work perceived as a replication, promoting the novelty of findings instead to avoid falling outside the scope or expectations of a conference and its reviewers. Authors may therefore tend to emphasize differences to related work over similarities, making replications and related results harder to identify and utilize.

In addition to these structural barriers, the HCS community may also find it particularly difficult to replicate studies because they often use scientific methodologies, such as qualitative methods, that are not designed to generate generalizable results and for which it is less clear what appropriate replication practices and goals would be, which is a topic of active discourse in other disciplines as well [28]. For example, exploratory interview studies are a commonly used qualitative study type in HCS, and while other researchers can repeat these studies using the interview guide, any differences with the original paper do not constitute a break with empiricism but rather add additional qualitative data points.

4. Discussion: Replication Crisis in HCS?

So, is HCS research in a replication crisis? To answer our initial question, we compare our findings to the replication crisis in psychology. Shrout and Rodgers describe three signals for the *replication crisis* in psychological research, which are also associated with the *"publish or perish*" pressure in academia [29]. The signals include:

1) *Scientific Fraudulent Behavior:* Cases of manipulated research data.
2) *Questionable Research Practices (QRP):* Cases of scientific fraud like p-hacking, HARKing (post-hoc hypothesizing), and selective reporting ("cherry picking" results).
3) *Failed Replications:* The Open Science Collaboration tried to replicate 100 top-tier published psychological studies in 2015, but only 36% of the results could be replicated, and effect sizes were on average half as large.

We discuss each indicator for the HCS community and share our view of where our community stands relative to them.

No documented fraudulent behavior (yet). Regarding the first signal, scientific fraudulent behavior, we are not aware of any documented cases of data fabrication or manipulation in the HCS community. While this is encouraging, it may also reflect the difficulty of detecting fraud in a field where raw data from user studies is rarely shared due to participant privacy constraints. Other fields, such as psychology, have been able to identify fraudulent studies by requiring the sharing of datasets alongside experimental findings [30].

Some questionable research practices. For the second signal, questionable research practices, there is reason for concern, but little concrete evidence. While Tang et al. report problematic statistic reporting in many human-centered security and privacy papers [31], practices such as p-hacking, HARKing, and selective reporting are difficult to detect from published papers alone. The transparency gaps documented by Klemmer et al. [5], [7] create conditions under which detecting questionable or unethical research practices is difficult. However, because systematic audits or pre-registered studies are uncommon in the field, we cannot confirm or rule out their prevalence.

Too few true replications to enable meta-analysis. The third signal, failed replications, presents an ambiguous picture. Of the 14 identified replication studies we analyzed (see Section 3.2), five confirmed previous results, four partially replicated them, and five produced significantly different findings. At first glance, a failure rate of roughly 40% among published replications might seem alarming. However, interpreting this number requires caution. The sample is small and likely not representative: many of the replication studies we found modified the original study design, population, or context. This makes it unclear whether divergent results indicate replication failures or reflect the intended study design variations. Moreover, interview participants in Klemmer et al.'s study reported that replications confirming original results are harder to publish [7], suggesting a potential publication bias that inflates the apparent rate of failed HCS replications when confirmed results are not considered novel enough by reviewers.

Flying blind. In summary, we cannot confidently state whether HCS research is in a replication crisis, given the lack of evidence. This in itself is problematic. The HCS field

lacks the infrastructure, incentives, and accumulated evidence to assess whether published results are independently verifiable. With only 14 labeled replication papers across more than two decades of HCS research, no systematic replication efforts akin to the Open Science Collaboration's work in psychology [29], and substantial transparency deficits [5] that impede replication attempts, the HCS community is effectively flying blind regarding the validity of its findings. The absence of evidence for a replication crisis should not be a comfort to the community, instead it should motivate the HCS community to look more closely. The structural barriers we have documented, including limited venue support for replications, missing study artifacts and intransparent reporting, and insufficient incentives to systematically conduct replications, suggest that the conditions for a crisis to develop undetected are firmly in place.

5. Calls to Action

Based on the current state of replications in HCS research and to avoid repeating the mistakes from other scientific disciplines, we see four areas that require attention by the community:

Standardizing transparent reporting. Transparency is a key necessity to replicating and validating scientific findings. To enable the replication of studies, the detection of scientific mistakes or wrongdoing, and the validation of findings across populations and study settings, providing study artifacts and core methodological information should be required for publications. The *"Transparency at SOUPS*" workshop at SOUPS 2025 [32] initiated a process to identify community guidelines for reporting study information, and their efforts resulted in a working document for transparent reporting [33]. However, while such guidelines are a necessary first step, we see this as an ongoing community-driven process to widely establish reporting recommendations and standards that enable replications and metascience.

Sharing research data. Research transparency ensures that study protocols can be replicated. Another key aspect for replicating research findings is access to the original study's dataset and analysis scripts. Without access to a full dataset, it is not possible to re-run the original analysis or to conduct comparative statistical analysis between the original and replicated datasets. Yet, there is reluctance within the HCS community to make datasets available due to concerns about participant privacy. However, most HCS studies produce data with limited confidentiality or disclosure risks, and techniques exist to further reduce disclosure risks in datasets, from basic removal of identifying information to statistical privacy measures, such as differential privacy. Furthermore, facilitating access to research data does not mean that a dataset must be either publicly available or not at all. The social sciences have established research data archives, such as ICPSR, that provide long-term, curated archival of research data, including management of restricted-use datasets [34]. Depending on a restricted-use dataset's sensitivity, access may require prior approval to download the dataset, be restricted to a virtual data enclave that prohibits data download, or be restricted to on-site physical access. The HCS community needs to set expectations for responsibly sharing research data, including appropriate participant privacy and dataset security measures.

Recognizing replications as valuable contributions. Replications are important, but as the results of Klemmer et al. and our CfP analysis suggest, they are difficult to publish at HCS venues, which deters researchers from conducting replications and leads to an over-emphasis on differences in findings across studies [5]. However, reporting and publishing confirmatory findings is equally important to enable meta-analysis of which HCS findings hold and generalize. Thus, replications need to be recognized and incentivized as valuable contributions by the community that should be encouraged at all HCS venues. Ideally, venues should encourage them prominently in their CfPs. In addition, reviewers must (be trained to) recognize the value of confirmatory findings for the scientific advancement of the security field.

Systematically guiding replication efforts. The few replication studies in HCS research we analyzed differ substantially in how they have been conducted, published, and labeled. It remains unclear under what conditions a study should be labeled as a replication and whether changes to the methodology and studied population are desirable for replication studies. To improve consistency and findability and to focus replication efforts on high community value, program chairs or venue steering committees could establish clearer guidelines and evaluation criteria that explicitly recognize the value of replication work. It remains an open question which studies are suited for replication and considered important and valuable to replicate, resulting in considerable insecurity for researchers who have to pick an *"important previously published study*" [8] to replicate. To identify prior results or research areas that could benefit from replication studies, venues with a clear HCS focus, like SOUPS or EuroUSEC could add Calls for Replications for specific studies or topics, e.g., authentication or security advice. Similarly, funding bodies, including government agencies and industry, could create research funding programs that promote replication of specific key findings or topics. Community competitions could also be established to promote replication studies.

References

[1] D. Evans. *"Nsf/iarpa/nsa workshop on the science of security*," Accessed: Feb. 17, 2026. [Online]. Available: https://sos.cs.virginia.edu/.

[2] D. Evans and S. Stolfo, *"Guest editors' introduction: The science of security*," *IEEE Security & Privacy*, vol. 9, no. 3, pp. 16–17, 2011.

[3] IEEE. *"31st ieee symposium on security and privacy — call for papers*," Accessed: Feb. 17, 2026. [Online]. Available: https://oakland31.cs.virginia.edu/oakland2010-cfp.pdf.

[4] C. Herley and P. C. van Oorschot, *"SoK: Science, Security, and the Elusive Goal of Security as a Scientific Pursuit*," in *Proc.\ 38th IEEE Symposium on Security and Privacy (SP'17)*, IEEE, 2017.

[5] J. H. Klemmer, J. Schmüser, B. M. Lowens, F. Fischer, L. Schmüser, F. Schaub, and S. Fahl, *"Transparency in usable privacy and security research: Scholars' perspectives, practices, and recommendations*," in *In 46th IEEE Symposium on Security and Privacy, IEEE S&P 2025, May 12-15, 2025*, IEEE, May 2025.

[6] ACM. *"Artifact review and badging*." version 1.1, Accessed: Feb. 13, 2026. [Online]. Available: https : / / www . acm . org / publications/policies/artifact-review-and-badging-current.

[7] J. H. Klemmer, J. Schmüser, F. Fischer, J. Suray, J.-U. Holtgrave, S. Lenau, B. M. Lowens, F. Schaub, and S. Fahl, *"How transparent is usable privacy and security research? a meta-study on current research transparency practices*," in *34th USENIX Security Symposium (USENIX Sec '25), August 13–15, 2025*, USENIX, Aug. 2025.

[8] Symposium On Usable Privacy and Security 2026. *"Call for papers*," Accessed: Feb. 18, 2026. [Online]. Available: https://soups.page/.

[9] K. Baig, E. Kazan, K. Hundlani, S. Maqsood, and S. Chiasson, *"Replication: Effects of media on the mental models of technical users*," in *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*, USENIX Association, Aug. 2021, pp. 119–138.

[10] A. Sotirakopoulos, K. Hawkey, and K. Beznosov, *"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*, Pittsburgh Pennsylvania: ACM, Jul. 20, 2011, pp. 1–18.

[11] E. Atwater, C. Bocovich, U. Hengartner, E. Lank, and I. Goldberg, *"Leading johnny to water: Designing for usability and trust*," in *Proceedings of the Eleventh USENIX Conference on Usable Privacy and Security*, ser. SOUPS '15, Ottawa, Canada: USENIX Association, 2015, pp. 69–88.

[12] S. Ruoti, N. Kim, B. Burgon, T. van der Horst, and K. Seamons, *"Confused johnny: When automatic encryption leads to confusion and mistakes*," in *Proceedings of the Ninth Symposium on Usable Privacy and Security*, ser. SOUPS '13, Newcastle, United Kingdom: Association for Computing Machinery, 2013.

[13] J.-U. Holtgrave, S. Klivan, K. Marky, and S. Fahl, *"A qualitative study of adoption barriers and challenges for passwordless authentication in german public administrations*," in *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, ser. CHI '25, Association for Computing Machinery, 2025.

[14] F. M. Farke, L. Lorenz, T. Schnitzler, P. Markert, and M. Dürmuth, *"{"you} still use the password after {all"}–exploring {fido2} security keys in a small company*," in *Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020)*, 2020, pp. 19–35.

[15] A. Danilova, A. Naiakshina, J. Deuter, and M. Smith, *"Replication: On the ecological validity of online security developer studies: Exploring deception in a password-storage study with freelancers*," in *Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020)*, USENIX Association, Aug. 2020, pp. 165–183.

[16] A. Naiakshina, A. Danilova, E. Gerlitz, E. von Zezschwitz, and M. Smith, *""if you want, i can store the encrypted password": A password-storage field study with freelance developers*," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19, Glasgow, Scotland Uk: Association for Computing Machinery, 2019, pp. 1–12.

[17] S. Zhuo, R. Biddle, J. Daniel Recomendable, G. Russello, and D. Lottridge, *"Eyes on the phish(er): Towards understanding users' email processing pattern and mental models in phishing detection*," in *Proceedings of the 2024 European Symposium on Usable Security*, ser. EuroUSEC '24, Association for Computing Machinery, 2024, pp. 15–29.

[18] C. Canfield, A. Davis, B. Fischhoff, A. Forget, S. Pearman, and J. Thomas, *"Replication: Challenges in using data logs to validate phishing detection ability metrics*," in *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*, Santa Clara, CA: USENIX Association, Jul. 2017, pp. 271–284.

[19] M. Volkamer, A. Gutmann, K. Renaud, P. Gerber, and P. Mayer, *"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)*, Baltimore, MD: USENIX Association, Aug. 2018, pp. 1–11.

[20] K. Busse, J. Schäfer, and M. Smith, *"Replication: No one can hack my mind revisiting a study on expert and non-expert security practices and advice*," in *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*, Santa Clara, CA: USENIX Association, Aug. 2019, pp. 117–136.

[21] S. Bird, I. Segall, and M. Lopatka, *"Replication: Why we still can't browse in peace: On the uniqueness and reidentifiability of web browsing histories*," in *Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020)*, USENIX Association, Aug. 2020, pp. 489–503.

[22] M. Schessler, E. Gerlitz, M. Häring, and M. Smith, *"Replication: Measuring user perceptions in smartphone security and privacy in germany*," in *Proceedings of the 2021 European Symposium on Usable Security*, 2021, pp. 165–179.

[23] A.-M. Ortloff, M. Vossen, and C. Tiefenau, *"Replicating a study of ransomware in germany*," in *Proceedings of the 2021 European Symposium on Usable Security*, 2021, pp. 151–164.

[24] J. Tang, E. Birrell, and A. Lerner, *"Replication: How well do my results generalize now? the external validity of online privacy and security surveys*," in *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*, Boston, MA: USENIX Association, Aug. 2022, pp. 367–385.

[25] K. Pfeffer, A. Mai, E. Weippl, E. Rader, and K. Krombholz, *"Replication: Stories as informal lessons about security*," in *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*, Boston, MA: USENIX Association, Aug. 2022, pp. 1–18.

[26] P. Kühtreiber, V. Pak, and D. Reinhardt, *"Replication: The effect of differential privacy communication on german users' comprehension and data sharing attitudes*," in *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*, Boston, MA: USENIX Association, Aug. 2022, pp. 117–134.

[27] A.-M. Ortloff, J. Tang, A. Arumugam, D. Huschina, L. Geierhaas, F. Martius, L. Jansen, K. Von Der Twer, L. Jungbluth, and M. Smith, *"Replication:{"no} one can hack my {mind"}-10 years later: An update and outlook on {experts'} and {non-experts'} security practices and advice*," in *Twenty-First Symposium on Usable Privacy and Security (SOUPS 2025)*, 2025, pp. 435–454.

[28] M. Pownall, *"Is replication possible in qualitative research? a response to makel et al. (2022)*," *Educational Research and Evaluation*, vol. 29, no. 1-2, pp. 104–110, 2024.

[29] P. E. Shrout and J. L. Rodgers, *"Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis*," *Annual review of psychology*, vol. 69, pp. 487–510, 2018.

[30] Association of Psychological Science. *"Psychological science submission guidelines*," Accessed: Feb. 18, 2026. [Online]. Available: https://www.psychologicalscience.org/publications/psychological_science/psci-submissions?utm_source=chatgpt.com.

[31] J. Tang, L. Bauer, and N. Christin, *"Misuse, misreporting, misinterpretation of statistical methods in usable privacy and security papers*," in *Proceedings of the Twenty-First USENIX Conference on Usable Privacy and Security*, ser. SOUPS '25, event-place: Seattle, WA, USA, USA: USENIX Association, 2025.

[32] S. Fahl, F. Schaub, J. H. Klemmer, J. Schmüser, B. M. Lowens, J.-U. Holtgrave, and J. Suray. *"Transparency at soups*," Accessed: Feb. 18, 2026. [Online]. Available: https://transparency-at-soups.teamusec.de/.

[33] J. H. Klemmer, J. Schmüser, J. Suray, J.-U. Holtgrave, B. Lowens, F. Schaub, and S. Fahl, *Transparency guidelines for human-centered security and privacy research*, Dec. 2025.

[34] ICPSR. *"Restricted-use data at icpsr*," Accessed: Feb. 19, 2026. [Online]. Available: https://www.icpsr.umich.edu/sites/icpsr/posts/shared/restricted-data-mgmt-access.