Meta-Science in Usable Security and Privacy and HCI

Anna-Marie Ortloff[1,*], Florin Martius[1]

[1]University of Bonn, Institute for Computer Science 4, Friedrich-Hirzebruch-Allee 5, 53115 Bonn, Germany

Abstract

Despite being a relatively young field, Usable Security and Privacy (USP) research has produced meta-research on methods, reporting practices, reproducibility, and evaluation. Much of this work focuses on study design and methodology, while broader research practices receive less attention. In this paper, we categorize existing meta-research in Usable Security and Privacy (USP) and examine how Usable Security and Privacy (USP) venues promote such efforts. We then present recommendations from our own meta-scientific studies on data handling, analysis, and reporting. Finally, we outline future directions to enhance the transparency, rigor, and impact of Usable Security and Privacy (USP) research.

Keywords

Usable security and privacy, meta-research, research methods

1. Introduction

Usable Security and Privacy (USP) research is a field at the intersection of IT security research and Human Computer Interaction (HCI), focusing on how humans interact with security technologies. This includes both end users [1], and expert users, such as software developers or administrators [2, 3]. There is a dedicated venue for USP research: the Symposium on Usable Privacy and Security (SOUPS), which has been taking place annually since 2005[1], but USP research is also often published at security conferences, mainly at USENIX Security, the IEEE Conference on Security and Privacy (S&P) and the ACM Conference on Computer and Communications Security (CCS), as well as at HCI venues like the ACM Conference on Human Factors in Computing Systems (CHI). Conducting studies in USP is especially challenging since security is often a secondary task for users, so the study design needs to take into account possible bias through priming participants about security or privacy and how to include real or simulated risks into study design [4, 5]. As USP researchers, we have encountered these and other challenges when conducting research in this domain. In this position paper, we first provide an overview of meta-science published with relation to USP and present a summary of meta-science as promoted or supported at venues where USP research is published. We then present recommendations for research practice derived from our own meta-scientific work in the areas of data handling, data analysis and reporting. Finally, we discuss potential directions for future meta-research in the USP domain.

2. Meta-science in Usable Security and Privacy

Ioannidis et al. identify five areas of meta-research: Research methods, reporting practice, reproducibility of research, evaluation methods, and incentivizing good science [6]. In the following, we present an overview of meta-scientific work in the USP community, along these five areas of meta-research. This is not intended to be a formal review, but a sample of meta-scientific work in USP and we omit our own work, which we discuss later. Le Pochat and Joosen reviewed meta-scientific work in security and privacy, including USP research next to more systems focused research [7]. We extend their work,

Meta-HCI: First Workshop on Meta-Research in HCI at ACM CHI'25, April 26, 2025, Yokohama, Japan
*Corresponding author.
ortloff@cs.uni-bonn.de (A. Ortloff); martius@cs.uni-bonn.de (F. Martius)
0000-0002-5735-178X (A. Ortloff); 0000-0001-8483-4042 (F. Martius)

[1]https://cups.cs.cmu.edu/soups/2005/SOUPS_2005_Conference_Report.html

narrowing our focus to USP rather than security and privacy more generally and explain where we think findings or methods from USP meta-research can be useful to the HCI community.

Most meta-scientific USP research focuses on methods, and several studies investigated aspects of study design. Danilova et al. [8] and Fulton et al. [9] compared security study tasks for developers – asking them to write, review, or fix code. In multiple studies, Naiakshina and colleagues explored priming effects, finding that priming significantly impacted secure password storage implementation [10, 11, 12, 13]. Deception is used in USP [10, 14], but also HCI [15]. Danilova et al. found that removing deception in a password storage study did not affect outcomes [16]. USP research also explored differences due to the study environment, either within a single publication [17] or a replication in a different publication [18] and identified differences between online and laboratory studies. Redmiles et al. found systematical differences between survey responses and field data, with respondents self-reporting faster software update speeds in surveys than observed in practice [19].

Some USP research requires access to hard-to-reach participants, such as software developers. Meta-scientific work has focused on remote recruitment, e.g. for crowd-sourced data collection [20, 21, 22], and identifying suitable recruitment channels [23, 24]. Serafini et al. [25, 26] focused on recruiting software developers working in companies and work by Naiakshina et al. compared different types of participants with programming skill: students, freelance developers and company developers [13], finding effect sizes differed but trends aligned. More broadly, studies have examined the representativeness of online recruitment for security studies, noting shifts over time [27, 28, 29]. A meta study found that even more participants in USP studies belong to Western, Educated, Industrialized, Rich, and Democratic (WEIRD) demographic groups than at CHI [30, 31].

USP meta-research on reporting is limited, though work by Groß [32] and Coopamootoo and Groß [33] found statistical reporting in USP is often incomplete. A poster publication on SOUPS future work statements showed limited impact on future research [34, 35]. Regarding replicability, as we discuss in Section 3, several USP venues explicitly support replications, leading to published replications across various areas [36, 16, 37, 28, 38, 39]. Hamm et al. [40] found study descriptions and materials are increasingly available at S&P, CCS, and USENIX Security, but data remain scarce..

Research on evaluating the research process is rare. Soneji et al. interviewed Program Committee members on peer review in security conferences [41]. While venues adapt submission and review processes, these changes rarely undergo scientific evaluation. Recent shifts include an artifact availability badge and review process, e.g. at USENIX Security in 2020[2] and a move toward journal-style submission cycles at security conferences, e.g. S&P, starting in 2018 with 12 submission dates per year, then quarterly in 2021 and then thrice yearly in 2022. We discuss further meta-science developments through analyzing calls for papers in Section 3. We were not able to find USP meta-research on incentivizing good science.

In our opinion, many USP meta-research findings, especially about methods, are at least partly transferable to HCI. For example, research on priming or recruiting software developers, though focused on security studies, can inform HCI research targeting the same demographic. It is necessary to reinvestigate issues outlined in descriptive meta-research [30, 40, 32], although for some topics, like participant demographics, the USP study [30] followed the methods of work about CHI [31].

3. Meta-science at Usable Security and Privacy Venues

In the following, we present an overview on how support for meta-scientific research is communicated in calls for papers at venues publishing USP papers, namely SOUPS, USENIX Security, CCS, and S&P, and compare our findings to CHI.

The concept of Systematization of Knowledge (SoK) papers was first introduced at S&P in 2010:

> In addition to the standard research papers, we are also soliciting papers this year focused on systematization of knowledge. The goal of this call is to encourage work that evaluates, systematizes, and contextualizes existing knowledge. These papers will provide

[2] https://www.usenix.org/conference/usenixsecurity20/artifact-evaluation-information

> a high value to our community but would otherwise not be accepted because they lack novel research contributions. Suitable papers include survey papers that provide useful perspectives on major research areas, papers that support or challenge long-held beliefs with compelling evidence, or papers that provide an extensive and realistic evaluation of competing approaches to solving specific problems. - *from the call for papers*[3]

This explicitly welcomes papers contributing to the areas of methods and reproducibility. Other venues also welcome SoK papers: SOUPS since 2019[4] and USENIX Security since 2024[5]. In the excerpt from the 2010 S&P call for papers above, replications are solicited as "papers that support or challenge long-held beliefs with compelling evidence", but not explicitly called replications. Several venues do explicitly include replications in their calls for papers. At SOUPS, replications have been named as one solicited topic since 2012[6], but have been additionally emphasized as a separate submission category since 2017[7]. At USENIX security, replication and reproduction of results were named as a subtopic of their newly introduced call for meta-scientific work in 2025:

> Meta-science, or the study of scientific research itself, aims to enhance the efficiency, quality, and outcomes of research activities in our community. Submissions in this broad topic should focus on evaluations of research practices, replicability/reproducibility, ethics, research methodologies, data transparency, and peer-review processes. Contributions should extend beyond analysis, aiming to influence future research practices.
>
> **Replication and Reproduction:** Contributions to this sub-topic should primarily consist of studies that verify, refute, or refine prior technical results or widely-held beliefs. We encourage submissions that not only replicate studies but also offer meta-analyses that assess the replicability of research. [...] - *from the call for papers*[8]

CCS does not solicit meta-science in their call for papers, neither as SoK contributions, nor as replications, or meta-science explicitly.

At CHI, neither SoK, nor meta-scientific contributions are explicitly mentioned in the call for papers, but replication has been named as a possible contribution in the guide for a successful submission since 2021[9]. The current subcommittees, of which authors of CHI submissions are required to choose at least one, are mostly very topically focused, i.e. *Games and Play* or indeed *Privacy and Security*. The subcommittees on *Understanding People*, using *Qualitative*, *Quantitative*, or *Mixed Methods*, represent somewhat of an exception, but their description still focuses on user studies, and does not explicitly welcome, for example, meta-analyses. This makes it hard to choose an appropriate subcommittee for researchers aiming to submit meta-scientific work to CHI. Since meta-science is relevant across different disciplines, we believe that the meta-scientific calls to action from USP venues are transferable to the HCI community.

4. Recommendations for the Research Process

Derived from our own meta-research at the intersection of USP and HCI, we present recommendations for researchers in three distinct areas of the research process: data handling, data analysis for qualitative research, which are both part of what Ioannidis et al. summarize as research focused on methods [6], and reporting on qualitative and quantitative analyses. Even though most of our work was conducted in the context of USP, many of these recommendations can be transferred to HCI, although additional

[3] https://www.ieee-security.org/TC/SP2010/archived/cfp.html
[4] https://www.usenix.org/conference/soups2019/call-for-papers
[5] https://www.usenix.org/conference/usenixsecurity24/call-for-papers
[6] https://cups.cs.cmu.edu/soups/2012/cfp.html
[7] https://www.usenix.org/conference/soups2017/call-for-papers
[8] https://www.usenix.org/conference/usenixsecurity25/call-for-papers
[9] https://web.archive.org/web/20210614064453/https://chi2021.acm.org/for-authors/presenting/papers/guide-to-a-successful-submission

verification for some results in a different research domain may be appropriate. Parts of this section are directly from prior published work.

4.1. Data handling

Ethical and legal standards require researchers to handle personal data responsibly, yet the realities of research workflows often make systematic data deletion difficult in practice. Data may be retained unnecessarily due to potential future use, unclear policies, or simple oversight. While researchers generally intend to protect participants' data, we believe structured processes can help ensure these intentions are realized. Based on our forthcoming research on data protection practices among USP researchers and insights from Data Protection Officers [42], we recommend the following:

Plan data retention and deletion from the beginning. Define which data must be kept and for how long before data collection begins, applying minimal data collection principles to reduce unnecessary retention.

Integrate clean-up into research workflows. Review and delete unnecessary data at key milestones, such as after analysis, publication, or project completion, to prevent accumulation.

Standardized templates and clear guidelines support responsible data management and should be fostered by all stakeholders. Reviewers should acknowledge ethical data deletion as best practice rather than penalizing missing data, ensuring that requests for additional data align with data minimization principles and do not encourage unnecessary retention.

4.2. Analysis of Qualitative Data

How qualitative data is coded – by a single researcher or a team – can influence the results. Multiple coders bring different perspectives but increase effort. Based on our research on coding outcomes across data types and coder groups, and a survey on reviewer expectations [43], we recommend:

Adapt coding practices to the complexity of research questions and data. For open-ended and complex research questions and data, use multiple coders. For simple data with well-defined research questions and straightforward analysis, a single coder can suffice if the codebook is well-defined. When using multiple coders, discussion and interaction are essential—avoid analyzing in isolation.

4.3. Reporting on Qualitative Analysis

To ensure transparency in reporting qualitative analysis, researchers should document both their methods and their expertise [43]:

Describe the involved researchers. Be clear about who the researchers doing the analysis are and what level of expertise they have, both in terms of the subject matter and the analysis methods used.

Describe analysis methods in detail, so that your process can be repeated and explain and justify important methodological decisions.

4.4. Reporting on Quantitative Analysis

Based on three meta-scientific publications, we make recommendations on reporting quantitative analyses, within the null hypothesis significance testing (NHST) framework. The first study manually extracted statistical data from developer-centered USP to support a-priori power analysis [44]. The second, forthcoming at CHI'25, used LLM-assisted extraction of statistical data from five years of quantitative CHI publications [45]. Finally, an interview and survey study with HCI and USP researchers, also forthcoming at CHI'25, examined researchers' interpretation and understanding of the odds ratio and Cohen's d effect size measures [46]. We have the following main recommendations for authors:

Report all effect sizes, even non significant ones. To gain a full understanding of the research area and reduce publication bias, non-significant effects are also important to report.

Report both standardized and non-standardized effect sizes. Our studies showed that researchers had trouble interpreting standardized effect sizes, particularly if they were unfamiliar with

the measure [46]. Thus we believe that simple effect-sizes in the units of the study should also always be reported, to aid interpretation. Both forms of effect-size have strengths and weaknesses, so offering both seems beneficial to us.

Interpret and discuss effect sizes. The interpretation should relate the results to the context of the research, including closely-related prior work, or a broader research area [45]. Interpretation should happen at an appropriate level of detail considering the importance of the discussed effect. For the main results of a study, practical relevance of the effect should be included in the discussion.

Explain effect size measures. In our studies, not all participants were familiar with the effect sizes we investigated, even though we had chosen ones which we believed to be common [46]. Thus, providing a short explanation can make the results more accessible to researchers not familiar with them.

Report complete and appropriate descriptive statistics. This includes reporting frequencies, when comparing nominal variables and reporting descriptive statistics for each group of the independent variables in the analysis, when comparing ordinal or interval variables. Such descriptives should be at least means, standard deviations and group sizes, where appropriate. Descriptive statistics should be applicable to the underlying data, e.g., a mean is not a good summary statistic for a bimodal distribution.

Make fully anonymized data sets available when needed. If presenting all the descriptive statistics and frequencies is too extensive, e.g., for regression analyses with multiple independent variables, ideally, the anonymized data can be made available.

Specify confirmatory or exploratory analysis. This helps assess stability of claims without hindering discovery [47].

We highlight two potentially controversial recommendations to encourage discussion:

Move test statistics to supplemental material. Test statistics (e.g. t, u, degrees of freedom) were not explicitly used the interviewed researchers when interpreting effects size [46], but they are relevant to verify statistical analyses. However, when providing complete information on statistical results, results sections may become cumbersome to read. As a middle ground between completeness of reporting and readability, these values, and detailed descriptive statistics could be reported in supplemental material, instead of in the main body of the paper, if they are not explicitly used for interpretation. Ideally this would be done in a machine-readable format.

Consider not reporting p-values. P-values are often the focus of interpretation of results [48, 46], even though they are actually more a measure of uncertainty. Using other ways to represent uncertainty, e.g. confidence intervals, could help researchers focus their interpretations more on effect sizes.

Finally, we have a recommendation and call to action for the HCI and USP communities:

Use reporting guidelines. In other fields, using guidelines, such as CONSORT [49], has been shown to improve reporting of study procedures [50]. Guidelines could help authors and reviewers judge and improve the completeness of statistical reporting, leading to improved ease in conducting meta-analyses, power analyses and building on prior work. Transparent statistics guidelines for HCI are in development[10], but are not complete and need to be extended.

5. Avenues for further research

Despite extensive research on the research process, much remains to be improved. As USP researchers, we believe "users are not the enemy" and should be supported in acting securely [51, 52]. We believe that this perspective also applies to researchers in USP and HCI, where usability-focused approaches can enhance research processes. Meta-science should go beyond identifying issues and actively support researchers by developing tools that implement best practices. These could aid confidential data handling through digital management plans and improve statistical analysis accessibility, helping both researchers and readers interpret results. With its expertise in user-friendly design, the HCI community is well-positioned to take on this challenge.

[10] https://transparentstats.github.io/guidelines/index.html

Declaration on Generative AI

During the preparation of this work, the author(s) used ChatGPT-4o and DeepL for grammar, spelling checks, and phrasing. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] A. Franz, V. Zimmermann, G. Albrecht, K. Hartwig, C. Reuter, A. Benlian, J. Vogt, {SoK}: Still Plenty of Phish in the Sea — A Taxonomy of {User-Oriented} Phishing Interventions and Avenues for Future Research, in: Proceedings of the Seventeenth Symposium on Usable Privacy and Security, USENIX Association, 2021, pp. 339–358.

[2] M. Tahaei, K. Vaniea, A Survey on Developer-Centred Security, in: 2019 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), IEEE, 2019, pp. 129–138. doi:10.1109/EuroSPW.2019.00021.

[3] C. Tiefenau, M. Häring, K. Krombholz, Security, Availability, and Multiple Information Sources: Exploring Update Behavior of System Administrators, in: Sixteenth Symposium on Usable Privacy and Security, USENIX Association, 2020, pp. 239–258.

[4] V. Distler, M. Fassl, H. Habib, K. Krombholz, G. Lenzini, C. Lallemand, L. F. Cranor, V. Koenig, A Systematic Literature Review of Empirical Methods and Risk Representation in Usable Privacy and Security Research, ACM Transactions on Computer-Human Interaction 28 (2021) 43:1–43:50. doi:10.1145/3469845.

[5] A. Sotirakopoulos, K. Hawkey, 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, Association for Computing Machinery, New York, NY, USA, 2011, pp. 1–18. doi:10.1145/2078827.2078831.

[6] J. P. A. Ioannidis, D. Fanelli, D. D. Dunne, S. N. Goodman, Meta-research: Evaluation and Improvement of Research Methods and Practices, PLOS Biology 13 (2015) e1002264. doi:10.1371/journal.pbio.1002264.

[7] V. Le Pochat, W. Joosen, Analyzing Cyber Security Research Practices through a Meta-Research Framework, in: Proceedings of the 16th Cyber Security Experimentation and Test Workshop, Association for Computing Machinery, New York, NY, USA, 2023, pp. 64–74. doi:10.1145/3607505.3607523.

[8] A. Danilova, A. Naiakshina, A. Rasgauski, M. Smith, Code reviewing as methodology for online security studies with developers - a case study with freelancers on password storage, in: Seventeenth Symposium on Usable Privacy and Security, USENIX Association, 2021, pp. 397–416.

[9] K. R. Fulton, J. Lewis, N. Malkin, M. L. Mazurek, Write, Read, or Fix? Exploring Alternative Methods for Secure Development Studies, in: Twentieth Symposium on Usable Privacy and Security, 2024, pp. 81–100.

[10] A. Naiakshina, A. Danilova, C. Tiefenau, M. Smith, Deception Task Design in Developer Password Studies: Exploring a Student Sample, in: Proceedings of the Fourteenth Symposium on Usable Privacy and Security, USENIX Association, 2018, pp. 297–313.

[11] A. Naiakshina, A. Danilova, C. Tiefenau, M. Herzog, S. Dechand, M. Smith, Why Do Developers Get Password Storage Wrong? A Qualitative Usability Study, in: Proceedings of the ACM SIGSAC Conference on Computer and Communications Security, Association for Computing Machinery, New York, NY, USA, 2017, pp. 311–328. doi:10.1145/3133956.3134082.

[12] A. Naiakshina, A. Danilova, E. Gerlitz, E. von Zezschwitz, M. Smith, "If you want, I can store the encrypted password": A Password-Storage Field Study with Freelance Developers, in: Proceedings of the CHI Conference on Human Factors in Computing Systems, ACM, Glasgow, Scotland, Uk, 2019, pp. 1–12. doi:10.1145/3290605.3300370.

[13] A. Naiakshina, A. Danilova, E. Gerlitz, M. Smith, On Conducting Security Developer Studies with

CS Students: Examining a Password-Storage Study with CS Students, Freelancers, and Company Developers, in: Proceedings of the CHI Conference on Human Factors in Computing Systems, ACM, Honolulu HI USA, 2020, pp. 1–13. doi:10.1145/3313831.3376791.

[14] V. Distler, The Influence of Context on Response to Spear-Phishing Attacks: An In-Situ Deception Study, in: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, ACM, Hamburg Germany, 2023, pp. 1–18. doi:10.1145/3544548.3581170.

[15] K. Katsuragawa, Q. Shu, E. Lank, PledgeWork: Online Volunteering through Crowdwork, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA, 2019, pp. 1–11. doi:10.1145/3290605.3300541.

[16] A. Danilova, A. Naiakshina, J. Deuter, M. Smith, Replication: On the Ecological Validity of Online Security Developer Studies: Exploring Deception in a {Password-Storage} Study with Freelancers, in: Proceedings of the Sixteenth Symposium on Usable Privacy and Security, USENIX Association, 2020, pp. 165–183.

[17] S. Fahl, M. Harbach, Y. Acar, M. Smith, On the ecological validity of a password study, in: Proceedings of the Ninth Symposium on Usable Privacy and Security, Association for Computing Machinery, New York, NY, USA, 2013, pp. 1–13. doi:10.1145/2501604.2501617.

[18] A. Sotirakopoulos, K. Hawkey, K. Beznosov, "I did it because I trusted you": Challenges with the Study Environment Biasing Participant Behaviours, in: SOUPS Usable Security Experiment Reports (USER) Workshop, ACM, 2010, p. 6.

[19] E. M. Redmiles, Z. Zhu, S. Kross, D. Kuchhal, T. Dumitras, M. L. Mazurek, Asking for a Friend: Evaluating Response Biases in Security User Studies, in: Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, ACM, Toronto Canada, 2018, pp. 1238–1255. doi:10.1145/3243734.3243740.

[20] A. Danilova, A. Naiakshina, S. Horstmann, M. Smith, Do you Really Code? Designing and Evaluating Screening Questions for Online Surveys with Programmers, in: IEEE/ACM 43rd International Conference on Software Engineering (ICSE), 2021, pp. 537–548. doi:10.1109/ICSE43902.2021.00057.

[21] A. Danilova, S. Horstmann, M. Smith, A. Naiakshina, Testing time limits in screener questions for online surveys with programmers, in: Proceedings of the 44th International Conference on Software Engineering, Association for Computing Machinery, New York, NY, USA, 2022, pp. 2080–2090. doi:10.1145/3510003.3510223.

[22] R. Serafini, C. Otto, S. A. Horstmann, A. Naiakshina, ChatGPT-Resistant Screening Instrument for Identifying Non-Programmers, in: Proceedings of the IEEE/ACM 46th International Conference on Software Engineering, Association for Computing Machinery, New York, NY, USA, 2024, pp. 1–13. doi:10.1145/3597503.3639075.

[23] H. Kaur, S. Amft, D. Votipka, Y. Acar, S. Fahl, Where to Recruit for Security Development Studies: Comparing Six Software Developer Samples, in: 31st USENIX Security Symposium, USENIX Association, Boston, MA, USA, 2022, p. 18.

[24] M. Tahaei, K. Vaniea, Recruiting Participants With Programming Skills: A Comparison of Four Crowdsourcing Platforms and a CS Student Mailing List, in: CHI Conference on Human Factors in Computing Systems, ACM, New Orleans LA USA, 2022, pp. 1–15. doi:10.1145/3491102.3501957.

[25] R. Serafini, M. Gutfleisch, S. A. Horstmann, A. Naiakshina, On the Recruitment of Company Developers for Security Studies: Results from a Qualitative Interview Study, in: Proceedings of the Nineteenth Symposium on Usable Privacy and Security, USENIX Association, Anaheim, USA, 2023, pp. 321–340.

[26] R. Serafini, S. A. Horstmann, A. Naiakshina, Engaging Company Developers in Security Research Studies: A Comprehensive Literature Review and Quantitative Survey, in: 33rd USENIX Security Symposium (USENIX Security 24), 2024, pp. 3277–3294.

[27] E. M. Redmiles, S. Kross, M. L. Mazurek, How Well Do My Results Generalize? Comparing Security and Privacy Survey Results from MTurk, Web, and Telephone Samples, in: 2019 IEEE Symposium on Security and Privacy (SP), 2019, pp. 1326–1343. doi:10.1109/SP.2019.00014.

[28] J. Tang, E. Birrell, A. Lerner, Replication: How Well Do My Results Generalize Now? The External

Validity of Online Privacy and Security Surveys, in: Proceedings of the Eighteenth Symposium on Usable Privacy and Security, USENIX Association, Boston, USA, 2022, pp. 367–385.
[29] R. Kang, S. Brown, L. Dabbish, S. Kiesler, Privacy Attitudes of Mechanical Turk Workers and the {U.S}. Public, in: 10th Symposium On Usable Privacy and Security (SOUPS 2014), USENIX Association, Menlo Park, USA, 2014, pp. 37–49.
[30] A. A. Hasegawa, D. Inoue, M. Akiyama, How WEIRD is Usable Privacy and Security Research?, in: 33rd USENIX Security Symposium (USENIX Security 24), USENIX Association, Philadelphia, USA, 2024, pp. 3241–3258.
[31] S. Linxen, C. Sturm, F. Brühlmann, V. Cassau, K. Opwis, K. Reinecke, How WEIRD is CHI?, in: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA, 2021, pp. 1–14. doi:10.1145/3411764.3445488.
[32] T. Groß, Fidelity of Statistical Reporting in 10 Years of Cyber Security User Studies, in: T. Groß, T. Tryfonas (Eds.), Socio-Technical Aspects in Security and Trust, Lecture Notes in Computer Science, Springer International Publishing, Cham, 2021, pp. 3–26. doi:10.1007/978-3-030-55958-8_1.
[33] K. Coopamootoo, T. Gross, A Systematic Evaluation of Evidence-Based Methods in Cyber Security User Studies, Technical Report CS_TR-1518, Newcastle University School of Computing, 2019.
[34] J. Suray, J. H. Klemmer, J. Schmüser, S. Fahl, How the Future Works at SOUPS: Analyzing Future Work Statements and Their Impact on Usable Security and Privacy Research, 2024. doi:10.48550/arXiv.2405.20785. arXiv:2405.20785.
[35] J. Suray, J. H. Klemmer, J. Schmüser, S. Fahl, Poster: Future Work Statements at SOUPS, in: Proceedings of the Twentieth Symposium on Usable Privacy and Security, USENIX Association, 2024.
[36] S. Bird, I. Segall, 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, 2020, pp. 489–503. URL: https://www.usenix.org/conference/soups2020/presentation/bird.
[37] K. Baig, E. Kazan, K. Hundlani, S. Maqsood, 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, 2021, pp. 119–138. URL: https://www.usenix.org/conference/soups2021/presentation/baig.
[38] P. Kühtreiber, V. Pak, 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), USENIX Association, Boston, MA, 2022, pp. 117–134. URL: https://www.usenix.org/conference/soups2022/presentation/kuhtreiber.
[39] K. Pfeffer, A. Mai, E. Weippl, E. Rader, K. Krombholz, Replication: Stories as informal lessons about security, in: Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022), USENIX Association, Boston, MA, 2022, pp. 1–18. URL: https://www.usenix.org/conference/soups2022/presentation/pfeffer.
[40] P. Hamm, D. Harborth, S. Pape, A Systematic Analysis of User Evaluations in Security Research, in: Proceedings of the 14th International Conference on Availability, Reliability and Security, Association for Computing Machinery, New York, NY, USA, 2019, pp. 1–7. doi:10.1145/3339252.3340339.
[41] A. Soneji, F. B. Kokulu, C. Rubio-Medrano, T. Bao, R. Wang, Y. Shoshitaishvili, A. Doupé, "Flawed, but like democracy we don't have a better system": The Experts' Insights on the Peer Review Process of Evaluating Security Papers, in: 2022 IEEE Symposium on Security and Privacy (SP), IEEE, San Francisco, CA, USA, 2022, pp. 1845–1862. doi:10.1109/SP46214.2022.9833581.
[42] F. Martius, L. Jansen, L. Struck, A. Arumugam, L. Geierhaas, A.-M. Ortloff, M. Smith, C. Tiefenau, Out of sight, out of mind? Exploring data protection practices for personal data in usable security & privacy studies, in: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, Yokohama, Japan, 2025. doi:10.1145/3706598.3713654.

[43] A.-M. Ortloff, M. Fassl, A. Ponticello, F. Martius, A. Mertens, K. Krombholz, M. Smith, Different Researchers, Different Results? Analyzing the Influence of Researcher Experience and Data Type During Qualitative Analysis of an Interview and Survey Study on Security Advice, in: Proceedings of the CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA, 2023, pp. 1–21. doi:10.1145/3544548.3580766.

[44] A.-M. Ortloff, C. Tiefenau, M. Smith, SoK: I have the (Developer) Power! Sample Size Estimation for Fisher's Exact, Chi-Squared, McNemar's, Wilcoxon Rank-Sum, Wilcoxon Signed-Rank and t-tests in Developer-Centered Usable Security, in: Proceedings of the Nineteenth Symposium on Usable Privacy and Security, USENIX Association, Anaheim, CA, 2023, pp. 341–359.

[45] A.-M. Ortloff, F. Martius, M. Meier, T. Raimbault, L. Geierhaas, M. Smith, Small, medium, large? A meta-study of effect sizes at CHI to aid interpretation of effect sizes and power calculation, in: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, Yokohama, Japan, 2025. doi:10.1145/3706598.3713671.

[46] A.-M. Ortloff, J. A. Grohs, S. Lenau, M. Smith, A qualitative study on how usable security and HCI researchers judge the size and importance of odds ratio and cohen's d effect sizes, in: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, 2025. doi:10.1145/3706598.3714022.

[47] A. T. Tredennick, G. Hooker, S. P. Ellner, P. B. Adler, A practical guide to selecting models for exploration, inference, and prediction in ecology, Ecology 102 (2021). doi:10.1002/ecy.3336.

[48] L. Besançon, P. Dragicevic, The Continued Prevalence of Dichotomous Inferences at CHI, in: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, ACM, Glasgow Scotland Uk, 2019, pp. 1–11. doi:10.1145/3290607.3310432.

[49] K. F. Schulz, D. G. Altman, D. Moher, CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials, Journal of Pharmacology and Pharmacotherapeutics 1 (2010) 100–107. doi:10.4103/0976-500X.72352.

[50] A. C. Plint, D. Moher, A. Morrison, K. Schulz, D. G. Altman, C. Hill, I. Gaboury, Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review, Medical Journal of Australia 185 (2006) 263–267. doi:10.5694/j.1326-5377.2006.tb00557.x.

[51] A. Adams, M. A. Sasse, Users are not the enemy, Communications of the ACM 42 (1999) 40–46. doi:10.1145/322796.322806.

[52] M. Green, M. Smith, Developers are Not the Enemy!: The Need for Usable Security APIs, IEEE Security & Privacy 14 (2016) 40–46. doi:10.1109/MSP.2016.111.