A Systematic Analysis of User Evaluations in Security Research

Peter Hamm
peter.hamm@m-chair.de
Goethe University Frankfurt
Frankfurt, Germany

David Harborth
david.harborth@m-chair.de
Goethe University Frankfurt
Frankfurt, Germany

Sebastian Pape
sebastian.pape@m-chair.de
Goethe University Frankfurt
Frankfurt, Germany

ABSTRACT

We conducted a literature survey on reproducibility and replicability of user surveys in security research. For that purpose, we examined all papers published over the last five years at three leading security research conferences and recorded the type of study and whether the authors made the underlying responses available as open data, as well as if they published the used questionnaire respectively interview guide. We uncovered how user surveys become more widespread in security research and how authors and conferences are increasingly publishing their methodologies, while we had no examples of data being made available. Based on these findings, we recommend that future researchers publish their data in addition to their results to facilitate replication and ensure a firm basis for user studies in security research.

CCS CONCEPTS

• **General and reference** → **Surveys and overviews**; *Empirical studies*; • **Security and privacy** → **Human and societal aspects of security and privacy**; *Social aspects of security and privacy*; *Usability in security and privacy*.

KEYWORDS

systematic literature review, qualitative methods, quantitative methods, user evaluations, human aspects of security

ACM Reference Format:
Peter Hamm, David Harborth, and Sebastian Pape. 2019. A Systematic Analysis of User Evaluations in Security Research. In *Proceedings of the 14th International Conference on Availability, Reliability and Security (ARES 2019) (ARES '19), August 26–29, 2019, Canterbury, United Kingdom.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3339252.3340339

1 INTRODUCTION

Recently, there is an ongoing discussion in security research [46] and science in general [7, 24, 34] about the reproducibility of experiments and the sharing of research data and code. In a Nature article from 2016 named "1,500 scientists lift the lid on reproducibility", 52% of the 1,576 surveyed researchers stated that they see a "significant reproducibility crisis" in their respective field [5]. Since then, several initiatives[1] and special issues in journals have been started to conduct replication research and promote higher standard with respect to reproducibility.

Madeyski et al. [46] consider reproducibility as a minimum requirement for good science. They define reproducible research as research papers that "incorporate the basic data and specific statistical methods used to analyze those data". Madeyski et al. [46] distinguish between replication of a study where independent researchers repeat the same experiment with different subjects, and thus generate a new dataset and reproducibility where the existing data set along with the used tools and the process described in the study should lead to the same results. There is also an ongoing discussion about the terminology of reproducibility and replicability [58] along with definitions from the International Vocabulary of Metrology [36] and the corresponding standard ISO 5725-2, the Association for Computing Machinery (ACM) [4] and others where both terms have different meanings. However, we will not elaborate in depth on the differences in terminology for the context of our study. The core of both concepts is transparency provided by researchers explaining their methodology and analysis in a detailed manner as well as clearly showing potential drawbacks and limitations of the respective approaches and results.

In general, when assessing the types of security research, we found:

Theoretical contributions where the authors developed a model or a framework or did a proof (e.g. cryptography)
→ similar to research in mathematics and philosophy

Construction of a system where the authors built a system
→ similar to research in computer science and (software) engineering

Measurements within a system where the authors (technically) evaluated a system (e.g. measure of throughput, performance, etc.)
→ similar to research in biology, physics and chemistry

User studies where the authors considered the users of a system
→ similar to research in psychology, sociology and economics

After that classification, we decided to investigate user studies in security research to analyze their reproducibility and replicability. We decided to focus on this type of studies since they become increasingly important as a component to evaluate the user value of technical solutions in the computer science discipline (cf. Figure 1 and 2). In light of the aforementioned replication issues in fields heavily relying on on surveys and experiments with individuals, the research community reacted and engages in efforts to replicate existing work for several years now, and latest studies indicate that

[1] For example, Center for Open Science https://cos.io

ARES '19, August 26–29, 2019, Canterbury, United Kingdom

ACM ISBN 978-1-4503-7164-3/19/08…$15.00
https://doi.org/10.1145/3339252.3340339

a relatively large share of social science studies is still not replicable (38%) [65]. Such studies with a focus on the user evaluations and assessments have not been conducted in the computer science discipline, to the best of our knowledge.

Following the definition of Madeyski et al. and considering the additional need of transparency, openness as promoted by Nosek et al. [53], we conclude that the availability of data along with a precise description of its evaluation is necessary to allow reproducibility and a rigorous description of the executed experiment.

Thus, as a first step, we conduct a literature survey on user studies in security research and investigate how many of them document the used method, rigorously report data collection and analysis, including listing the questionnaire items and interview guides as well as which studies make the research data available.

Literature reviews or surveys can be conducted following several different methodologies. In order to have reproducible and well documented results, we have decided to follow the methodology for conducting literature reviews by vom Brocke et al. [74], combined with a concept-centric approach for synthesizing the results [76]. Due to the massive amount of security research today, the literature review covers only a representative selection of three high quality general security conferences for the search process. This is a regular limitation for systematic literature reviews since it is close to impossible to cover all existing outlets of a discipline. The remainder of the paper is as follows. Section 2 discussed related work before Section 3 describes the methodology in detail. Section 4 presents the results of the literature review and the analysis and synthesis. Section 5 discusses the results and concludes with recommendations for future user studies in the security field.

2 RELATED WORK

Related work with the goal of investigating the degree of replicability and reproducibility of scientific research gained substantial traction in the last years. This is mainly due to the revelations related to faked or tweaked research data and results and the following reproducibility crisis [5]. Thus, several researchers set out to lay more focus on this important issue. This is done by either providing special outlets for such papers or by conducting large-scale studies with the goal to replicate existing research.

An example for a new journal, specifically focusing on replication work, is the Transaction of Replication Research (TRR) in the Information Systems (IS) domain [21]. Here, authors must submit either exact replications, methodological replications or conceptual replications and thereby define the exact type of replication study they conduct.

Examples for large-scale studies investigating multiple existing research articles can be found in several disciplines. One of the most criticized disciplines with respect to reproducibility is psychology. Here, a large-scale replication study investigating 100 experimental and correlational studies found that the mean effect size found in the replication studies (0.197) was half the effect size found in the original papers (0.403) [56]. Besides that, several effects were roughly zero in the replication paper indicating that the results found in the original papers must be questioned with respect to their reliability.

Related to our field of research is a study by Collberg et al. [14, 15] who did a study on papers from computer systems researchers and investigated the extent to which they share their code and data and to the extent to which the received code builds.

They examined 402 papers from ACM conferences and journals and were able to obtain the code for 32% of them and build it within 30 minutes. For 48% (54%) they were able to build the code with extra effort (or the authors stated it would build with reasonable effort). They suggest a sharing specification scheme to specify the level of sharing for a paper.

3 METHODOLOGY

The literature review follows the framework by vom Brocke et al. [74]. This framework suggests building a taxonomy for the literature review in order to provide a comprehensible structure (cf. Table 1).

The grey cells show the applicable categories of this literature review. The review aims for research methods with respect to user studies and evaluations applied in security research articles. The goal is to integrate the used methods of the research articles to provide a common ground for future researchers and to identify guidelines for conducting quantitative user studies in the security domain. The structure of the results is supposed to be methodological with a neutral representation of the results. The review aims at the audience of scholars specialized in the field of security research. The coverage strives for representative results with respect to the top three security conferences (cf. Table 2). We decided to focus on conferences rather than journals as we believe that conferences have a tendency to offer more current results and thus better capture new trends, while journal articles tend to undergo longer review and publication cycles and may potentially lag behind conferences in terms of very new developments.

The selection of the included conferences is based on the Core Conference Ranking [16]. We included only the general security conferences ranked A+ in this ranking. Following this selection criterion, we included the IEEE Symposium on Security and Privacy, the ACM Computer and Communications Security Conference and the USENIX Security Symposium. In addition, we only searched the main conference proceedings (i.e. no workshops or comparable co-located events), as we felt that this would ensure an unbiased look at the relative consideration that user studies are receiving in this field. We include the proceedings of the conferences ranging from 2013 onwards. To follow the goal of this literature review and provide a profound and diverse insight into current user studies in security research, we did not use any keywords to search the three conferences, but checked each paper in every year in the respective proceedings. The sum of all papers in the conference proceedings are regarded as "Hits". Subsequently, the hits are evaluated thoroughly based on whether they employ a quantitative method and conduct user studies. If a paper does this, it is considered as a "Final Hit" and the methods are analyzed in-depth.

We defined user-study as research aiming at understanding the user through self-reported answers, be it qualitatively through interviews or quantitatively through surveys. As an example, we decided to exclude Bonneau and Schechter [10], who would distract their users and measure how well they recall their passwords, as this

Table 1: Taxonomy of the Literature Review following Cooper [17]

Characteristic	Categories			
Focus	research outcomes	research methods	theories	applications
Goal	integration	criticism	central issues	
Organization	historical	conceptual	methodological	
Perspective	neutral representation		espousal of position	
Audience	specialized scholars	general scholars	practitioners/politicians	general public
Coverage	exhaustive	exhaustive and selective	representative	central/pivotal

was measured directly through the input speed of the users, rather than relying on the users answering questions.

The structure of the literature search is based on the proposal by vom Brocke et al. [74]. On the top level, we distinguish between qualitative research, which is defined by Strauss and Corbin [70, p. 10f] as: "[…] any type of research that produces findings not arrived at by statistical procedures or other means of quantification", and quantitative research which makes use of statistical procedures. Generally, if an article incorporated both qualitative and quantitative surveys and one is given significantly more weight than another, we generally assigned the paper to the category receiving more focus. Only if a publication incorporates both substantial surveys with large sample sizes as well as detailed interviews that go beyond pretesting for the surveys did we refer to is as mixed methods. The article by Egelman et al. [23] is a representative example for these kinds of studies including structured interviews in combination with a large user-study.

Finally, we checked whether the authors offered the survey or interview responses they obtained as open data for future research and whether they added the specific questions the participants were asked.

Table 2 shows our findings from the literature search. Following our selection criteria, we could identify 61 articles which are in the scope of our literature review (marked as "Final Hits"). We found most user-related studies in the USENIX Security Symposium (26) followed by the ACM CSS with 19 articles and the IEEE S&P with 16 articles. Figure 1 illustrates the number of articles over our period of coverage. It can be seen that there is a slight overall increase in user studies for all three conference together with a major increase in 2018.

4 RESULTS

Table 3 presents the results of the literature search. Overall, we see a notable absolute as well as relative increase in the proportion of user studies compared to other studies as evidenced by Figures 1 and 2, with only 2016 slightly bucking this trend on the back of an unusually high number of articles published in the USENIX Security Symposium in 2015, whose subsequent fall in the next year could not be counteracted by a rise in the number of user studies in the other two conferences. Furthermore, even though none of the authors offered the full data from which they derived their results, Table 3 and Figure 3 show that 31 out of 61 papers supplied their questionnaires or interview guides, which provides immense value for replications with different participants. A general trend is that the proportion of supplied questionnaires increased in time; all nine

Figure 1: Overall number of user studies since 2013

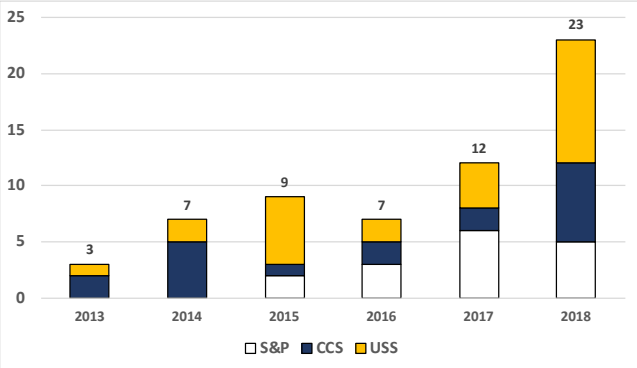

Figure 2: Proportion of user studies since 2013

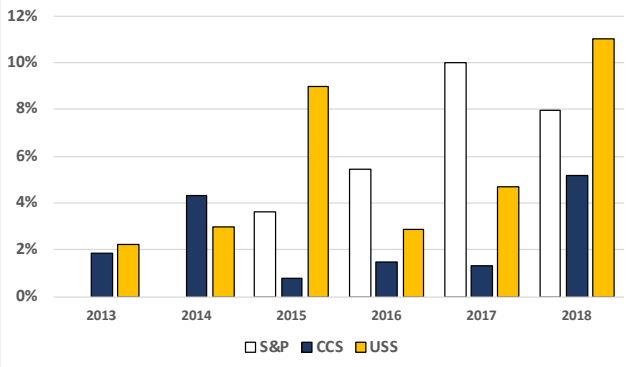

Figure 3: Relative number of papers supplying questions

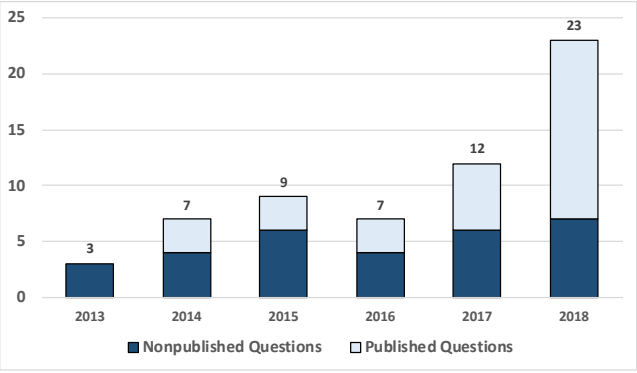

Table 2: Summary of the Literature Search

	Conference	Coverage	Hits	Final Hits
1	IEEE Symposium on Security and Privacy	2013 – 2018	315	16
2	ACM Computer and Communications Security Conference	2013 – 2018	1,050	19
3	USENIX Security Symposium	2013 – 2018	437	26
Sum			1,802	61

Table 3: Results of the Literature Review following the structure proposed by vom Brocke et al. [74]

		Quantitative Method		Qualitative Method		Mixed Method
		Descriptive	Inferential	given	not given	
Experimental design	Field		Komanduri et al. [39] Plane et al. [57][2] Harkous et al. [32][2] Bianchi et al. [8] Tischer et al. [72][2] Joudaki et al. [37][2] Golla et al. [31][2] Redmiles et al. [62][2] Nguyen et al. [52][2] Das et al. [18][2]	Gallagher et al. [29][2]		Stevens et al. [69]
	Lab	Fratantonio et al. [28]	Karapanos et al. [38][2] Tian et al. [71][2] Acar et al. [3] Lévesque et al. [43]	Lebeck et al. [42][2] Naiakshina et al. [50][2]	Lau et al. [41] Reynolds et al. [63][2]	Krombholz et al. [40][2]
Non-experimental design		Jana et al. [35] Feng et al. [27] Fawaz et al. [25] Li et al. [44] Yuan et al. [80] Vilk et al. [73] Cherubini et al. [12][2] Sahin and Francillon [64] Neupane et al. [51] Polakis et al. [59][2] Fawaz and Shin [26] Shirvanian and Saxena [67] Denning et al. [20]	Dechand et al. [19][2] Lyastani et al. [45][2] Wijesekera et al. [78] Olejnik et al. [54] Acar et al. [2] Cho et al. [13] Sharif et al. [66][2] Redmiles et al. [60]	McGregor et al. [47][2] Redmiles et al. [61][2] Abu-Salma et al. [1] Votipka et al. [75][2] Blond et al. [9][2] Simko et al. [68]	Oltrogge et al. [55] Becker et al. [6] Mu et al. [49][2]	Wijesekera et al. [77][2] McGregor et al. [48][2] Winter et al. [79] Zhang-Kennedy et al. [82] Gao et al. [30][2] He et al. [33][2] Zeng et al. [81] Egelman et al. [23][23] Dietrich et al. [22][24]

[1] Open Data [2] Questionnaire reported [3] Egelman et al. [23] had structured interviews and a large user-study [4] Dietrich et al. [22] had a (non-saturated) set of interviews as well as a quantitative study

user studies presented at the ACM Computer and Communications Security Conference since 2017 supplied their questionnaires or interview guides, compared to four out of ten in the period between 2013 and 2016. The same indications are observable for the other two conferences we consider. We also observe a higher proportion of supplied questions for studies employing inferential statistics compared to those that focus on purely descriptive methods, which is most likely indicative that the user surveys played a higher role for those studies, while descriptive statistics are generally being offered as supplements to studies less focused on user opinions. The proportion of studies employing qualitative methods that supplied their interview guides was not significantly different from the corresponding proportion for studies that supplied questionnaires.

5 DISCUSSION AND CONCLUSION

Our results point to a growing role of user studies in security research, which is still dominated by more technical approaches. This shift underscores the importance of supplying detailed methodological information by the authors of said studies to ensure replicability, especially considering the warnings of a replicability crisis in social sciences [11], which rely heavily on equivalent methodologies. Our review shows encouraging signs in this direction, with a growing proportion of authors supplying their questionnaires or survey guides; notably, each identified user study published at the ACM Computer and Communications Security Conference since 2016 had their questionnaires and study guides included in the appendix. Considering open data, our results were disillusioning. None of

the papers provided their data which would not only allow reproducibility, but also related studies to do a more accurate comparison of the results. We want to encourage openness for all conferences in security research, to ease replicability and stop spurious results from taking hold. Encouraging the addition of questionnaires to the appendices of user studies would be simple to achieve and can help put these results on firmer ground.

Future work could expand the literature surveyed. While we decided to focus on a small set of highly ranked conferences, additional conferences as well as journals may give a more detailed look into the state of users surveys in security research. In addition to a wider breadth, future research could also look deeper into specific methods employed in the user studies. We decided to focus on a high-level overview and not to differentiate between different statistical tests, even though we saw a diversity of approaches. We also omitted a deeper look into sample sizes, as there was not always a clear distinction between preliminary tests and studies, and the first group could easily distort results. For qualitative studies, there were structured interviews [23], semi-structured interviews [50], and group discussions [68], but we did not dig deep into their theoretical foundations lest we distract from the big picture. Future work could further investigate the current use of pretests in user studies in security research. We have seen surveys used as pretests for semistructured interviews [50], as well as surveys being conducted following non-saturated interviews [22], and this is an interesting phenomenon that may warrant further consideration.

ACKNOWLEDGMENTS

This work received funding from the European Union's Horizon 2020 research and innovation programme from the project CyberSec4Europe (grant agreement number 830929).

REFERENCES

[1] Ruba Abu-Salma, M. Angela Sasse, Joseph Bonneau, Anastasia Danilova, Alena Naiakshina, and Matthew Smith. 2017. Obstacles to the Adoption of Secure Communication Tools. In *IEEE Symposium on Security and Privacy*. IEEE Computer Society, 137–153.

[2] Yasemin Acar, Michael Backes, Sascha Fahl, Simson L. Garfinkel, Doowon Kim, Michelle L. Mazurek, and Christian Stransky. 2017. Comparing the Usability of Cryptographic APIs. In *IEEE Symposium on Security and Privacy*. IEEE Computer Society, 154–171.

[3] Yasemin Acar, Michael Backes, Sascha Fahl, Doowon Kim, Michelle L. Mazurek, and Christian Stransky. 2016. You Get Where You're Looking for: The Impact of Information Sources on Code Security. In *IEEE Symposium on Security and Privacy*. IEEE Computer Society, 289–305.

[4] Association for Computing Machinery. 2016. Artifact Review and Badging. Available online at: https://www.acm.org/publications/policies/artifact-review-badging.

[5] Monya Baker. 2016. 1,500 scientists lift the lid on reproducibility. https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970.

[6] Ingolf Becker, Simon Parkin, and M. Angela Sasse. 2018. The Rewards and Costs of Stronger Passwords in a University: Linking Password Lifetime to Strength. In *USENIX Security Symposium*. USENIX Association, 239–253.

[7] Jeremy Berg. 2018. Progress on reproducibility.

[8] Antonio Bianchi, Jacopo Corbetta, Luca Invernizzi, Yanick Fratantonio, Christopher Kruegel, and Giovanni Vigna. 2015. What the App is That? Deception and Countermeasures in the Android User Interface. In *IEEE Symposium on Security and Privacy*. IEEE Computer Society, 931–948.

[9] Stevens Le Blond, Alejandro Cuevas, Juan Ramón Troncoso-Pastoriza, Philipp Jovanovic, Bryan Ford, and Jean-Pierre Hubaux. 2018. On Enforcing the Digital Immunity of a Large Humanitarian Organization. In *IEEE Symposium on Security and Privacy*. IEEE Computer Society, 424–440.

[10] Joseph Bonneau and Stuart E. Schechter. 2014. Towards Reliable Storage of 56-bit Secrets in Human Memory. In *USENIX Security Symposium*. USENIX Association, 607–623.

[11] Colin F Camerer, Anna Dreber, Felix Holzmeister, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Gideon Nave, Brian A Nosek, Thomas Pfeiffer, et al. 2018. Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour* 2, 9 (2018), 637.

[12] Mauro Cherubini, Alexandre Meylan, Bertil Chapuis, Mathias Humbert, Igor Bilogrevic, and Kévin Huguenin. 2018. Towards Usable Checksums: Automating the Integrity Verification of Web Downloads for the Masses. In *ACM Conference on Computer and Communications Security*. ACM, 1256–1271.

[13] Geumhwan Cho, Jun Ho Huh, Junsung Cho, Seongyeol Oh, Youngbae Song, and Hyoungshick Kim. 2017. SysPal: System-Guided Pattern Locks for Android. In *IEEE Symposium on Security and Privacy*. IEEE Computer Society, 338–356.

[14] Christian Collberg, Todd Proebsting, and Alex M Warren. 2015. Repeatability and benefaction in computer systems research. *University of Arizona TR* 14 (2015), 4.

[15] Christian Collberg and Todd A Proebsting. 2016. Repeatability in computer systems research. *Commun. ACM* 59, 3 (2016), 62–69.

[16] Computing Research & Education. 2019. Core Conference Ranking. http://portal.core.edu.au/conf-ranks/.

[17] Harris M. Cooper. 1988. Organizing Knowledge Synthesis: A Taxonomy of Literature Reviews. *Knowledge in Society* 1 (1988), 104–126. https://doi.org/10.1007/BF03177550

[18] Sauvik Das, Adam D. I. Kramer, Laura A. Dabbish, and Jason I. Hong. 2014. Increasing Security Sensitivity With Social Proof: A Large-Scale Experimental Confirmation. In *ACM Conference on Computer and Communications Security*. ACM, 739–749.

[19] Sergej Dechand, Dominik Schürmann, Karoline Busse, Yasemin Acar, Sascha Fahl, and Matthew Smith. 2016. An Empirical Study of Textual Key-Fingerprint Representations. In *USENIX Security Symposium*. USENIX Association, 193–208.

[20] Tamara Denning, Adam Lerner, Adam Shostack, and Tadayoshi Kohno. 2013. Control-Alt-Hack: the design and evaluation of a card game for computer security awareness and education. In *ACM Conference on Computer and Communications Security*. ACM, 915–928.

[21] Alan R. Dennis and Joseph S. Valacich. 2014. A Replication Manifesto. *Transactions on Replication Research* 1, 1 (2014), 1–5.

[22] Constanze Dietrich, Katharina Krombholz, Kevin Borgolte, and Tobias Fiebig. 2018. Investigating System Operators' Perspective on Security Misconfigurations. In *ACM Conference on Computer and Communications Security*. ACM, 1272–1289.

[23] Serge Egelman, Sakshi Jain, Rebecca S. Portnoff, Kerwell Liao, Sunny Consolvo, and David A. Wagner. 2014. Are You Ready to Lock?. In *ACM Conference on Computer and Communications Security*. ACM, 750–761.

[24] Daniele Fanelli. 2018. Opinion: Is science really facing a reproducibility crisis, and do we need it to? *Proceedings of the National Academy of Sciences* 115, 11 (2018), 2628–2631.

[25] Kassem Fawaz, Huan Feng, and Kang G. Shin. 2015. Anatomization and Protection of Mobile Apps' Location Privacy Threats. In *USENIX Security Symposium*. USENIX Association, 753–768.

[26] Kassem Fawaz and Kang G. Shin. 2014. Location Privacy Protection for Smartphone Users. In *ACM Conference on Computer and Communications Security*. ACM, 239–250.

[27] Huan Feng, Kassem Fawaz, and Kang G. Shin. 2015. LinkDroid: Reducing Unregulated Aggregation of App Usage Behaviors. In *USENIX Security Symposium*. USENIX Association, 769–783.

[28] Yanick Fratantonio, Chenxiong Qian, Simon P. Chung, and Wenke Lee. 2017. Cloak and Dagger: From Two Permissions to Complete Control of the UI Feedback Loop. In *IEEE Symposium on Security and Privacy*. IEEE Computer Society, 1041–1057.

[29] Kevin Gallagher, Sameer Patil, Brendan Dolan-Gavitt, Damon McCoy, and Nasir D. Memon. 2018. Peeling the Onion's User Experience Layer: Examining Naturalistic Use of the Tor Browser. In *ACM Conference on Computer and Communications Security*. ACM, 1290–1305.

[30] Xianyi Gao, Yulong Yang, Can Liu, Christos Mitropoulos, Janne Lindqvist, and Antti Oulasvirta. 2018. Forgetting of Passwords: Ecological Theory and Data. In *USENIX Security Symposium*. USENIX Association, 221–238.

[31] Maximilian Golla, Miranda Wei, Juliette Hainline, Lydia Filipe, Markus Dürmuth, Elissa M. Redmiles, and Blase Ur. 2018. "What was that site doing with my Facebook password?": Designing Password-Reuse Notifications. In *ACM Conference on Computer and Communications Security*. ACM, 1549–1566.

[32] Hamza Harkous, Kassem Fawaz, Rémi Lebret, Florian Schaub, Kang G. Shin, and Karl Aberer. 2018. Polisis: Automated Analysis and Presentation of Privacy Policies Using Deep Learning. In *USENIX Security Symposium*. USENIX Association, 531–548.

[33] Weijia He, Maximilian Golla, Roshni Padhi, Jordan Ofek, Markus Dürmuth, Earlence Fernandes, and Blase Ur. 2018. Rethinking Access Control and Authentication for the Home Internet of Things (IoT). In *USENIX Security Symposium*. USENIX Association, 255–272.

[34] Matthew Hutson. 2018. Artificial intelligence faces reproducibility crisis.

[35] Suman Jana, David Molnar, Alexander Moshchuk, Alan M. Dunn, Benjamin Livshits, Helen J. Wang, and Eyal Ofek. 2013. Enabling Fine-Grained Permissions for Augmented Reality Applications with Recognizers. In *USENIX Security Symposium*. USENIX Association, 415–430.

[36] Joint Committee for Guides in Metrology. 2006. International Vocabulary of Metrology - Basic and General Concepts and Associated Terms. Available online at: https://www.nist.gov/sites/default/files/documents/pml/div688/grp40/International-Vocabulary-of-Metrology.pdf. 3rd Edition.

[37] Zeinab Joudaki, Julie Thorpe, and Miguel Vargas Martin. 2018. Reinforcing System-Assigned Passphrases Through Implicit Learning. In *ACM Conference on Computer and Communications Security*. ACM, 1533–1548.

[38] Nikolaos Karapanos, Claudio Marforio, Claudio Soriente, and Srdjan Capkun. 2015. Sound-Proof: Usable Two-Factor Authentication Based on Ambient Sound. In *USENIX Security Symposium*. USENIX Association, 483–498.

[39] Saranga Komanduri, Richard Shay, Lorrie Faith Cranor, Cormac Herley, and Stuart E. Schechter. 2014. Telepathwords: Preventing Weak Passwords by Reading Users' Minds. In *USENIX Security Symposium*. USENIX Association, 591–606.

[40] Katharina Krombholz, Wilfried Mayer, Martin Schmiedecker, and Edgar R. Weippl. 2017. "I Have No Idea What I'm Doing" - On the Usability of Deploying HTTPS. In *USENIX Security Symposium*. USENIX Association, 1339–1356.

[41] Billy Lau, Simon P. Chung, Chengyu Song, Yeongjin Jang, Wenke Lee, and Alexandra Boldyreva. 2014. Mimesis Aegis: A Mimicry Privacy Shield-A System's Approach to Data Privacy on Public Cloud. In *USENIX Security Symposium*. USENIX Association, 33–48.

[42] Kiron Lebeck, Kimberly Ruth, Tadayoshi Kohno, and Franziska Roesner. 2018. Towards Security and Privacy for Multi-user Augmented Reality: Foundations with End Users. In *IEEE Symposium on Security and Privacy*. IEEE Computer Society, 392–408.

[43] Fanny Lalonde Lévesque, Jude Nsiempba, José M. Fernandez, Sonia Chiasson, and Anil Somayaji. 2013. A clinical study of risk factors related to malware infections. In *ACM Conference on Computer and Communications Security*. ACM, 97–108.

[44] Frank Li, Zakir Durumeric, Jakub Czyz, Mohammad Karami, Michael Bailey, Damon McCoy, Stefan Savage, and Vern Paxson. 2016. You've Got Vulnerability: Exploring Effective Vulnerability Notifications. In *USENIX Security Symposium*. USENIX Association, 1033–1050.

[45] Sanam Ghorbani Lyastani, Michael Schilling, Sascha Fahl, Michael Backes, and Sven Bugiel. 2018. Better managed than memorized? Studying the Impact of Managers on Password Strength and Reuse. In *USENIX Security Symposium*. USENIX Association, 203–220.

[46] Lech Madeyski, Barbara A Kitchenham, and Shari Lawrence Pfleeger. 2015. Why Reproducible Research is Beneficial for Security Research. *under review* (2015).

[47] Susan E. McGregor, Polina Charters, Tobin Holliday, and Franziska Roesner. 2015. Investigating the Computer Security Practices and Needs of Journalists. In *USENIX Security Symposium*. USENIX Association, 399–414.

[48] Susan E. McGregor, Elizabeth Anne Watkins, Mahdi Nasrullah Al-Ameen, Kelly Caine, and Franziska Roesner. 2017. When the Weakest Link is Strong: Secure Collaboration in the Case of the Panama Papers. In *USENIX Security Symposium*. USENIX Association, 505–522.

[49] Dongliang Mu, Alejandro Cuevas, Limin Yang, Hang Hu, Xinyu Xing, Bing Mao, and Gang Wang. 2018. Understanding the Reproducibility of Crowd-reported Security Vulnerabilities. In *USENIX Security Symposium*. USENIX Association, 919–936.

[50] Alena Naiakshina, Anastasia Danilova, Christian Tiefenau, Marco Herzog, Sergej Dechand, and Matthew Smith. 2017. Why Do Developers Get Password Storage Wrong?: A Qualitative Usability Study. In *ACM Conference on Computer and Communications Security*. ACM, 311–328.

[51] Ajaya Neupane, Md. Lutfor Rahman, Nitesh Saxena, and Leanne M. Hirshfield. 2015. A Multi-Modal Neuro-Physiological Study of Phishing Detection and Malware Warnings. In *ACM Conference on Computer and Communications Security*. ACM, 479–491.

[52] Duc-Cuong Nguyen, Dominik Wermke, Yasemin Acar, Michael Backes, Charles Weir, and Sascha Fahl. 2017. A Stitch in Time: Supporting Android Developers in WritingSecure Code. In *ACM Conference on Computer and Communications Security*. ACM, 1065–1077.

[53] Brian A Nosek, George Alter, George C Banks, Denny Borsboom, Sara D Bowman, Steven J Breckler, Stuart Buck, Christopher D Chambers, Gilbert Chin, Garret Christensen, et al. 2015. Promoting an open research culture. *Science* 348, 6242 (2015), 1422–1425.

[54] Katarzyna Olejnik, Italo Dacosta, Joana Soares Machado, Kévin Huguenin, Mohammad Emtiyaz Khan, and Jean-Pierre Hubaux. 2017. SmarPer: Context-Aware and Automatic Runtime-Permissions for Mobile Devices. In *IEEE Symposium on Security and Privacy*. IEEE Computer Society, 1058–1076.

[55] Marten Oltrogge, Yasemin Acar, Sergej Dechand, Matthew Smith, and Sascha Fahl. 2015. To Pin or Not to Pin-Helping App Developers Bullet Proof Their TLS Connections. In *USENIX Security Symposium*. USENIX Association, 239–254.

[56] Open Science Collaboration. 2015. Estimating the reproducibility of psychological science. *Science* 349, 6251 (2015), aac4716. https://doi.org/10.1126/science.aac4716

[57] Angelisa C. Plane, Elissa M. Redmiles, Michelle L. Mazurek, and Michael Carl Tschantz. 2017. Exploring User Perceptions of Discrimination in Online Targeted Advertising. In *USENIX Security Symposium*. USENIX Association, 935–951.

[58] Hans E Plesser. 2018. Reproducibility vs. replicability: a brief history of a confused terminology. *Frontiers in neuroinformatics* 11 (2018), 76.

[59] Iasonas Polakis, Panagiotis Ilia, Federico Maggi, Marco Lancini, Georgios Kontaxis, Stefano Zanero, Sotiris Ioannidis, and Angelos D. Keromytis. 2014. Faces in the Distorting Mirror: Revisiting Photo-based Social Authentication. In *ACM Conference on Computer and Communications Security*. ACM, 501–512.

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

[61] Elissa M. Redmiles, Amelia R. Malone, and Michelle L. Mazurek. 2016. I Think They're Trying to Tell Me Something: Advice Sources and Selection for Digital Security. In *IEEE Symposium on Security and Privacy*. IEEE Computer Society, 272–288.

[62] Elissa M. Redmiles, Ziyun Zhu, Sean Kross, Dhruv Kuchhal, Tudor Dumitras, and Michelle L. Mazurek. 2018. Asking for a Friend: Evaluating Response Biases in Security User Studies. In *ACM Conference on Computer and Communications Security*. ACM, 1238–1255.

[63] Joshua Reynolds, Trevor Smith, Ken Reese, Luke Dickinson, Scott Ruoti, and Kent E. Seamons. 2018. A Tale of Two Studies: The Best and Worst of YubiKey Usability. In *IEEE Symposium on Security and Privacy*. IEEE Computer Society, 872–888.

[64] Merve Sahin and Aurélien Francillon. 2016. Over-The-Top Bypass: Study of a Recent Telephony Fraud. In *ACM Conference on Computer and Communications Security*. ACM, 1106–1117.

[65] Kelly Servick. 2018. 'Generous' approach to replication confirms many high-profile social science findings. https://www.sciencemag.org/news/2018/08/generous-approach-replication-confirms-many-high-profile-social-science-findings.

[66] Mahmood Sharif, Jumpei Urakawa, Nicolas Christin, Ayumu Kubota, and Akira Yamada. 2018. Predicting Impending Exposure to Malicious Content from User Behavior. In *ACM Conference on Computer and Communications Security*. ACM, 1487–1501.

[67] Maliheh Shirvanian and Nitesh Saxena. 2014. Wiretapping via Mimicry: Short Voice Imitation Man-in-the-Middle Attacks on Crypto Phones. In *ACM Conference on Computer and Communications Security*. ACM, 868–879.

[68] Lucy Simko, Ada Lerner, Samia Ibtasam, Franziska Roesner, and Tadayoshi Kohno. 2018. Computer Security and Privacy for Refugees in the United States. In *IEEE Symposium on Security and Privacy*. IEEE Computer Society, 409–423.

[69] Rock Stevens, Daniel Votipka, Elissa M. Redmiles, Colin Ahern, Patrick Sweeney, and Michelle L. Mazurek. 2018. The Battle for New York: A Case Study of Applied Digital Threat Modeling at the Enterprise Level. In *USENIX Security Symposium*. USENIX Association, 621–637.

[70] Anselm Strauss and Juliet Corbin. 1998. *Basics of qualitative research techniques*. Sage publications Thousand Oaks, CA.

[71] Yuan Tian, Nan Zhang, Yue-Hsun Lin, XiaoFeng Wang, Blase Ur, Xianzheng Guo, and Patrick Tague. 2017. SmartAuth: User-Centered Authorization for the Internet of Things. In *USENIX Security Symposium*. USENIX Association, 361–378.

[72] Matthew Tischer, Zakir Durumeric, Sam Foster, Sunny Duan, Alec Mori, Elie Bursztein, and Michael Bailey. 2016. Users Really Do Plug in USB Drives They Find. In *IEEE Symposium on Security and Privacy*. IEEE Computer Society, 306–319.

[73] John Vilk, David Molnar, Benjamin Livshits, Eyal Ofek, Christopher J. Rossbach, Alexander Moshchuk, Helen J. Wang, and Ran Gal. 2015. SurroundWeb: Mitigating Privacy Concerns in a 3D Web Browser. In *IEEE Symposium on Security and Privacy*. IEEE Computer Society, 431–446.

[74] Jan vom Brocke, Alexander Simons, Bjoern Niehaves, Kai Riemer, Ralf Plattfaut, and Anne Cleven. 2009. Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. In *ECIS 2009 Proceedings*. 2206–2217. arXiv:ECIS2009-0566.R1

[75] Daniel Votipka, Rock Stevens, Elissa M. Redmiles, Jeremy Hu, and Michelle L. Mazurek. 2018. Hackers vs. Testers: A Comparison of Software Vulnerability Discovery Processes. In *IEEE Symposium on Security and Privacy*. IEEE Computer Society, 374–391.

[76] Jane Webster and Richard T Watson. 2002. Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MISQ Quarterly* 26, 2 (2002), xiii–xxiii.

[77] Primal Wijesekera, Arjun Baokar, Ashkan Hosseini, Serge Egelman, David A. Wagner, and Konstantin Beznosov. 2015. Android Permissions Remystified: A Field Study on Contextual Integrity. In *USENIX Security Symposium*. USENIX Association, 499–514.

[78] Primal Wijesekera, Arjun Baokar, Lynn Tsai, Joel Reardon, Serge Egelman, David A. Wagner, and Konstantin Beznosov. 2017. The Feasibility of Dynamically Granted Permissions: Aligning Mobile Privacy with User Preferences. In *IEEE Symposium on Security and Privacy*. IEEE Computer Society, 1077–1093.

[79] Philipp Winter, Anne Edmundson, Laura M. Roberts, Agnieszka Dutkowska-Zuk, Marshini Chetty, and Nick Feamster. 2018. How Do Tor Users Interact With Onion Services?. In *USENIX Security Symposium*. USENIX Association, 411–428.

[80] Xuejing Yuan, Yuxuan Chen, Yue Zhao, Yunhui Long, Xiaokang Liu, Kai Chen, Shengzhi Zhang, Heqing Huang, Xiaofeng Wang, and Carl A. Gunter. 2018. CommanderSong: A Systematic Approach for Practical Adversarial Voice Recognition. In *USENIX Security Symposium*. USENIX Association, 49–64.

[81] Kexiong (Curtis) Zeng, Shinan Liu, Yuanchao Shu, Dong Wang, Haoyu Li, Yanzhi Dou, Gang Wang, and Yaling Yang. 2018. All Your GPS Are Belong To Us: Towards Stealthy Manipulation of Road Navigation Systems. In *USENIX Security Symposium*. USENIX Association, 1527–1544.

[82] Leah Zhang-Kennedy, Hala Assal, Jessica N. Rocheleau, Reham Mohamed, Khadija Baig, and Sonia Chiasson. 2018. The aftermath of a crypto-ransomware attack at a large academic institution. In *USENIX Security Symposium*. USENIX Association, 1061–1078.