

# Transparency in Usable Privacy and Security Research: Scholars’ Perspectives, Practices, and Recommendations

Jan H. Klemmer<sup>✉\*</sup>, Juliane Schmäser<sup>✉\*</sup>, Byron M. Lowens<sup>✉†</sup>, Fabian Fischer<sup>✉\*</sup>,  
Lea Schmäser<sup>‡</sup>, Florian Schaub<sup>✉†</sup>, and Sascha Fahl<sup>✉\*</sup>

<sup>\*</sup>*CISPA Helmholtz Center for Information Security, Germany,*

*{jan.klemmer, juliane.schmueser, fabian.fischer, sascha.fahl}@cispa.de*

<sup>†</sup>*University of Michigan, USA, {bmlowens, fschaub}@umich.edu*

<sup>‡</sup>*Leibniz University Hannover, Germany, lea.schmueser@sec.uni-hannover.de*

**Abstract**—Transparent reporting of research is a crucial aspect of good scientific practice and contributes to trustworthy science. Transparency helps to understand research processes, assess the validity of research contributions, and facilitates replication of studies and reported results. In the face of reproducibility crises in other fields, the security and privacy (SP) research community in general and the usable privacy and security (UPS) community in particular lack clear standards for transparent research reporting.

To gain insights into current research transparency practices and associated challenges and obstacles in the UPS community, we report findings from 24 semi-structured interviews with UPS researchers. We find that researchers value research transparency and already apply several transparency reporting practices. However, an implicit community standard without incentives that outweigh challenges and drawbacks appears to prevent further advances in research transparency. Based on our findings, we conclude with recommendations for transparency practices and guidance for publication venues to better incentivize research transparency (e.g., adapting artifact evaluation to typical UPS artifacts like study materials) and to alleviate constraints that hinder transparency (e.g., removing page limits on appendices). We hope our findings can spur community discussion and effort to improve research quality through more transparent research reporting.

## 1. Introduction

The transparent reporting of research is essential for trustworthy science. In the past, SP research has been criticized for lacking sound scientific procedures [1]. The SP community responded with an effort to improve the *Science of Security (SoS)* [2]. Amongst others, improved replicability, reproducibility, and transparent research reporting are essential goals of the SoS effort. Research disciplines, including psychology, social sciences, and biology, have been reported to be in *replication crises* [3]—while other researchers doubt this claim [4], [5]. It is consensus, though, that a lack of transparent research reporting significantly contributes to replication issues [4]. Transparency is also necessary for high-quality peer reviewing. Reviewers can

only assess the merit of submitted research when relevant methodological details are provided. Replicability and reproducibility require all relevant information and materials. We define the relevant terms in Section 2.

In 2017, Herley and van Oorschot illustrated the need for more transparent research reporting in security [2]. In response, multiple venues introduced artifact evaluations (AEs) (e.g., USENIX Security [6] and ACM CCS [7]). Recent results suggest a limited impact, as the availability of artifacts did not significantly improve [8]. We lack a deeper understanding of practices and experiences regarding transparent research reporting in the SP community. This is further complicated by the broader SP community’s heterogeneity, which includes many subfields—ranging from theory, systems, web, network, and artificial intelligence to human-centered privacy and security. Each subfield has transparency requirements, e.g., detailed proofs in theory, software artifacts in systems security, or study materials, measurement tools, and data analysis details in user and usability studies.

Since addressing all subfields in one study would lack research depth, in this paper, we focus on understanding the challenges of transparent research reporting in one subfield: *usable privacy and security (UPS) research*.<sup>1</sup> UPS research is diverse in itself due to various research methods, ranging from contextual inquiry and interviews to usability evaluations, controlled experiments, and measurement studies. Furthermore, UPS research often involves human subjects and requires careful consideration of ethical implications. Schechter’s opinion piece [9] from 2010 provides tips on avoiding common pitfalls when writing UPS papers. However, the UPS community lacks explicit guidelines and established transparency practices based on community consensus. While calls for papers (CfPs) and AE procedures encourage transparency, researcher practices and perspectives are highly individual and implicit.

We aim to shed light on transparency practices and perspectives of UPS researchers. To gain insights on current

1. We use the term *usable privacy and security (UPS)* to refer to the broader field of human-centered security and privacy, not just usability. We consider studies to be UPS if they focus on human factors or usability within security and privacy, typically human subjects research.

transparency practices and perspectives in UPS research, we conducted 24 semi-structured interviews with UPS researchers, representing a cross-section of the field including experienced researchers, community leaders, industry researchers, postdocs, and PhD students. Our interviews provide novel in-depth insights on these research questions:

**RQ1:** *What are UPS researchers’ practices regarding research transparency?*

**RQ2:** *What benefits, issues, and concerns do UPS researchers associate with research transparency?*

**RQ3:** *How do UPS researchers perceive the current state of transparency? How has it, and how should it develop?*

In this work, we make the following contributions:

*Qualitative Insights on Transparency in UPS:* In the first comprehensive analysis on research transparency practices in the UPS community, as part of the broader SP community, we contribute novel insights on UPS researchers’ practices, experiences, and challenges, identifying six themes that complement insights from other academic fields.

*State of Transparency in UPS:* We found varied understandings of transparency, and an implicit transparency standard in the UPS community. We identify a set of typical transparency practices in UPS and a mismatch of incentives and drawbacks that negatively affect research transparency. Participants reported improvements in the field but desired further progress and an adaptation of AE to UPS.

*Transparency Recommendations:* We provide actionable recommendations for venues that publish UPS research and the UPS community at large. These include formalizing guidelines based on community consensus, providing transparency incentives (e.g., transparency awards), addressing challenges (e.g., page limits, artifact hosting, and participant privacy), and adding transparency checks to review forms.

## 2. Background

To set the context for this paper and clarify some often confused terms, we provide the following definitions and background discussions on transparency, replicability, and reproducibility.

### 2.1. Transparency

For research and its reporting, we define transparency as follows:

**Transparency:** Reporting all relevant details, especially methodological details and artifacts, needed to (1) assess the validity of a study and its results and (2) independently re-run reported studies.

This definition is based on the US Academy of Sciences’s definition’s to share “*details about their research, including study design, materials used, details of the system under study, operationalization of variables, measurement techniques, uncertainties in measurement in the system under study, and how data were collected and analyzed*” [4].

### 2.2. Replicability & Reproducibility

The previous definition reflects that transparency is related to replication and reproduction of results, as a prerequisite for both [10]. While only related to transparency, we discuss replicability and reproducibility here as these concepts provide necessary background for our study and were also brought up by participants.

Different and even contradicting definitions for replication and reproducibility exist [11], [12]. Therefore, the terms are not consistently used, if distinguished at all. The confusion is further supported by changes in the definitions. For example, the ACM swapped the meaning of *Results Reproduced* and *Results Replicated* in 2020 [13] between their AE guidelines v1.0 [14] and v1.1 [15] to align with other definitions. We adopt the current (v1.1) ACM definition for this paper, as the ACM is relevant for the community and the ACM policies are already used within many AEs. The ACM definitions [15] can be summarized as follows:

**Repeatability:** The *same team* can obtain the same results with the *same method* (e.g., repeating computations).

**Reproducibility:** A *different team* can obtain the same result with the *same method* (e.g., using original authors’ software artifacts).

**Replicability:** A *different team* can obtain the same result with a *different method* (e.g., using self-developed software artifacts).

Besides these general definitions, an important and debatable question is when the “same result” was obtained, e.g., in what margin of error—and whether replicability and reproducibility can apply to UPS-typical human subjects research. While it might be feasible to obtain the exact same results for computational experiments where one just needs to run the same software artifact on the study data, or highly controlled experiments with human participants, this is not directly applicable to the kind of human subjects studies typical for UPS. For example, reproducing the exact participant statements in an interview study is impossible, even when asking the same participants. Nonetheless, the concepts can be transferred to UPS research. SOUPS explicitly encourages replication studies in the CfP [16] and has published them [17]–[20] for several years. We note that the SOUPS definition is broader and differs from the ACM’s, as the 2024 CfP [16] states: “*Replications may follow the same protocol as the original study, or may vary one or more key variables to see whether the result is extensible (e.g., re-running a study with a sample from a different population).*”

Overall, successfully reproducing or replicating results has a nuanced meaning, depending on study type and academic field. The ACM AE policy underlines the nuances for different study types and especially the applicability to human subject studies:

*“It is easy to see how research articles that develop algorithms or software systems could be labeled as [replicated or reproduced]. Here, the artifacts could be implementations of algorithms or complete software systems, and replication would involve exercise of software, typically software provided by the author.*”

However, we intend these [AE] badges to be applicable to other types of research as well. For example, artifacts associated with human-subject studies [...] might be the collected data, as well as the scripts developed to analyze the data. “Replication” might focus on a careful inspection of the experimental protocol along with independent analysis of the collected data.” — [15]

For example, there exist concepts for replicating qualitative studies [21], [22], like replicating certain aspects from an earlier qualitative study to extend and validate its results, e.g., in different contexts.

### 3. Related Work

We discuss related work on (i) transparency and reproducibility meta-studies as well as (ii) transparency practices and guidelines, and highlight our study’s contributions.

#### 3.1. Transparency & Reproducibility Meta-Studies

Recent years have seen a movement for open and transparent research. Researchers declared reproducibility crises in fields such as medicine [23] or psychology [24]. Claims reach as far as expecting more research results to be false than true [25]. In response to these issues, the investment into reproducibility efforts is rising, e.g., at the US National Science Foundation (NSF) [26]. Ideas to enable and incentivize transparency include gamification [27] and interactive statistical result reporting [28]. Others raised concerns regarding privacy risks from published data and advancements in deanonymization techniques [29], and ethical and methodological implications of secondary data usage [30].

In computer science, literature reviews show that studies lack information required for transparency and reproducibility, e.g., in text mining studies [31], machine learning publications [8], [32]–[34], information system journals [35], and software engineering (SE) studies [36], [37]. Research publications lacked critical data sets, algorithm specifications, or experiment code [8], [31]. Meta-studies also identified a lack of bias discussion [38], missing statistical information like effect sizes [37], and a general lack of documentation of execution details [33], [34], even in publications that make an effort to be transparent [36]. SE researchers emphasized the criticality of replication studies to validate research results while also acknowledging that these are typically hard to publish [39], and have transparency problems of their own [40]. Meta-studies in the human–computer interaction (HCI) community illustrated that important journals lack comprehensive transparency guidelines [41]. Only a few ACM CHI papers from 2016 and 2017 openly released software artifacts [42]. Reviewing ACM CHI 2017 and 2022 publications suggests an improvement of reporting transparency in some areas, e.g., sharing interview guides, but no difference in other areas, like the sharing of software artifacts [43]. An analysis of risk representation in UPS publications uncovered a tendency of missing method details [44]. A recent review of 715 UPS papers found

that papers regularly do not report participant demographics, especially at non-HCI and non-UPS venues [45].

Besides literature reviews, surveys and interviews with researchers from diverse fields provide another perspective on reproducibility issues [46]–[48]. Researchers struggled with reproducing the research of others [3], [49], and uncovered a lack of motivation and incentives, high effort, and technical difficulties as barriers to transparent research reporting [50], [51]. In computer security, interviews on the peer-reviewing process revealed that reviewers consider missing details to be a negative aspect, possibly leading to rejection of a paper [52]. Surveying ACM CHI authors on their material and data sharing practices found that sharing is uncommon, citing difficulties with protecting personal identifiable information (PII), reliable hosting, and author motivation and resources [53].

We add to this body of research in-depth insights from 24 interviews on practices and challenges regarding transparent research reporting in the UPS field, specifically.

#### 3.2. Transparency Practices and Guidelines

Different research communities have developed practices and publishing guidelines to improve transparent research reporting and reproducibility [5], [54]. Originating from medicine, the *EQUATOR Network* provides guidelines on reporting different types of studies to improve accuracy and transparency [55]–[57], such as the *CONSORT* guidelines for randomized trials [58] and *SRQR* recommendations for reporting qualitative research [59]. Grant agencies, including the NIH in the US, set certain expectations for transparent research reporting [60]. The open science paradigm is related to transparency. In 2015, the *Transparency and Openness Promotion Guidelines (TOP)* were established to promote transparency, openness, and reproducibility [61], [62]. TOPs consists of eight modular standards that venues can adopt as a policy for their publications. The 2016 *FAIR principles* for scientific data management propose that data needs to be findable, accessible, interoperable, and reusable (FAIR) [63], [64].

The SE community provides guidance to conduct [65] and report experiments to ensure availability of information [66], [67] and support replicability [68]. In 2020, the *ACM SIGSOFT* presented *Empirical Standards for Software Engineering Research* checklists that set method-specific expectations for conducting and reporting research in software engineering [69], [70]. The HCI community supports transparent research reporting as well. In 2018, extended abstracts published at ACM CHI discussed the adoption of TOPs [71] and transparent statistic reporting [72]. The aforementioned literature review by Salehzadeh Niksirat *et al.* can serve as a guideline for ACM CHI papers [43]. In the computer security community, artifact evaluation are a recent effort to improve reproducibility and transparency. In 2024, NDSS, ACM CCS, and USENIX Security call for artifacts and offer (optional) AE. The *Security Research Artifacts* project [73] centrally provides instructions, results, and links to research artifacts in security publications.

These guidelines show the increasing criticality of transparent research reporting across fields. For UPS research, specific guidelines are rare so far. In 2010, Schechter wrote an article on how to write a SOUPS paper [9]. Besides, more recent reporting guidelines for user studies in the context of risk representation towards participants emerged [44]. However, our participants were widely unaware of these guidelines and practices. Therefore, we explore UPS researchers' experiences, challenges, and practices and conclude with UPS-specific recommendations to promote transparency.

## 4. Methodology

We conducted 24 semi-structured interviews with UPS researchers between June and October 2023 via Zoom in English. We describe our interview guide, recruitment, demographics, and qualitative analysis. We also discuss ethics and limitations and provide a positionality statement. We provide our materials online (see Availability section).

### 4.1. Interview Guide Development & Piloting

We developed the interview guide based on our research questions and refined it through four pilot interviews with researchers from our groups. We incorporated their feedback to enhance question clarity, added sub-questions, and reordered questions for a better interview flow. During the main interviews, we refined the guide based on participant feedback, making only minor adjustments to sub-questions as needed. After conducting 14 interviews, we expanded the interview guide with additional questions to explore emerging themes. The recorded main interview part (excluding intro and outro) lasted 01:04:09 hours on average (median 01:01:27). See our replication package (Availability section) for the interview guide.

### 4.2. Interview Structure

Our interviews consisted of five sections. Figure 1 illustrates the interview flow. All interview sections started with an open question to allow the interviewees to express their thoughts freely.

**Intro.** Each interview started with a brief introduction about the project and the interview procedure. We obtained consent for the audio recording from all interviewees. We informed them that we are interested in their personal opinions and would not judge their practices regarding research transparency. We then asked interviewees to introduce themselves and their research as a warm-up.

**Transparency Impressions.** This first block of questions explores impressions and expectations related to transparency in UPS research. This included interviewees' definition of transparency, criteria they deem important for a research publication to be transparent, and transparency requirements they are familiar with for venues they publish at. To foster an in-depth exploration, we did not provide our definition from Section 2, to not prime participants and

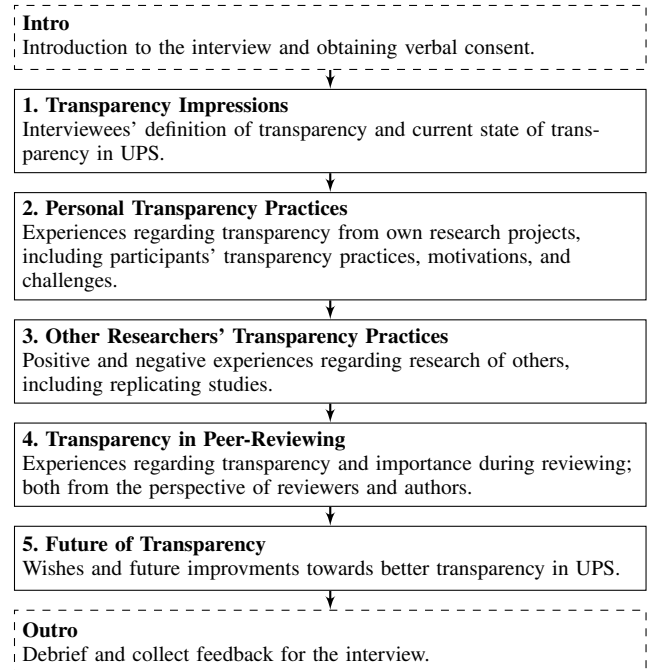


Figure 1. Structure of the semi-structured interviews. In each section, the interviewer asked general questions and corresponding follow-ups. Interviewees were generally free to diverge from this flow.

let them freely elaborate on the meaning of transparency. Moreover, we asked them to assess the current state of transparency in UPS research.

**Personal Transparency Practices.** Next, we elicited the interviewees' practices related to and motivation for reporting their research transparently. We started by asking how interviewees aim to be transparent. Furthermore, we prompted for advantages, disadvantages, and challenges interviewees experienced related to reporting research transparently. After that, to explore specific challenges, we prompted about the impact of page limits, data protection and intellectual property, ethical concerns, and technical limitations on transparency. Moreover, we explored interviewees' opinions on providing supplementary materials and artifact evaluation.

**Other Researchers' Transparency Practices.** Next, we focused on the participants' transparency experiences with the research and publications of other UPS researchers. In particular, we were interested in factors that positively or negatively impact transparency of research reporting. Additionally, we asked about experiences when replicating the research work of others (e.g., needed but missing materials or information).

**Transparency in Peer-Reviewing.** Then we asked about the role of transparent research reporting in the peer-review process. Our focus was on whether interviewees considered appendices and supplementary materials during reviews, their reasons for doing so, and if this consideration should be mandatory for all reviewers. Additionally, we delved into their experiences with reviews they had received, particularly focusing on how reviewers handle transparency

issues, their significance, and their influence on the review outcomes.

**Future of Transparency.** The last section explored potential improvements and desired changes related to transparent research reporting in the UPS community. This also included information and materials that papers should report for transparency and changes to CfPs.

**Additional Follow-Up Questions.** After the first 14 interviews, we added and asked additional questions before ending the interview. The goal was to follow-up on findings that emerged in the prior interviews, which is a common practice in semi-structured interviews [74], [75]. This included questions about replication, null results, pre-registration, registered reports, effort–benefit trade-offs, and incentives.

**Outro.** Finally, we asked for feedback, whether there were additional aspects interviewees wanted to mention, and answering questions from interviewees. We also asked them to fill out a demographic questionnaire (Availability section) and a form to indicate their compensation preferences.

### 4.3. Recruitment & Inclusion Criteria

We systematically recruited a diverse sample of 24 UPS researchers, including only participants who (i) published at least two UPS conference papers between 2018 and 2022, or (ii) served as program committee (PC) member or chair for SOUPS as the most relevant UPS venue. We began with six long-term community members who have published many UPS papers. We further randomly sampled 11 participants from the authors of UPS conference papers between 2018 and 2022 weighted by their number of publications. We identified UPS conference papers through manual coding of conference proceedings. The process is described in detail in Appendix A. To increase diversity, we added two PhD students and one postdoctoral researcher and recruited four researchers working in industry or for government agencies from the list of UPS paper authors by cross-referencing affiliations and personal websites. We excluded members of our research groups and close collaborators, and did not recruit multiple participants from the same research group. Our strategy yielded a sample that represents the community well, including recognized leaders, and junior and senior scholars from the community. We stopped after 24 interviews, having reached thematic saturation with no new themes and insights emerging after the analysis of interview 17 [76].

Senior team members who knew a potential participant sent recruitment emails to increase response rates. Team members without close ties to the participants conducted the interviews. Out of 37 contacted researchers, nine did not respond, and four declined due to time constraints or organizational policies. We offered participants a compensation of \$25 (or equivalent in local currency) as an Amazon voucher or PayPal payment. As recent studies with researchers had no consistent compensation approach [52], [77], we decided to make the compensation optional. Overall, 16 of 24 interviewees accepted the compensation.

### 4.4. Participant Demographics

To protect interviewee privacy within the small, well-connected UPS community, we report aggregated demographics. Our sample included 24 UPS researchers from the USA, Germany, and the UK, with a broad range of research experience: 13 with over 10 years of experiences (five of those had over 20 years), 10 with 6–10 years, and one with five years or less. Participant roles varied, including nine full, four associate, and three assistant professors, two PhD students, three research scientists/engineers, and one postdoc. One participant held a permanent academic non-faculty position; one chose not to disclose their role. Most participants had extensive reviewer experience, having served as PC or editorial board members (23), external (sub)reviewers (21), and PC or subcommittee/area chairs or editors (17). Methodological expertise was diverse. All participants had experience with multiple methods, with most having conducted interviews and surveys/questionnaires (20 each), experiments (18), measurement studies (15), literature reviews (14), and focus groups (11). The interviewees collectively have published 186 UPS papers between 2018 and 2022 ( $Md = 7.5$ ).

### 4.5. Qualitative Analysis Approach

We transcribed the audio recordings with the GDPR-compliant manual transcription service of *Amberscript* [78]. During analysis, we additionally reviewed the transcripts for any transcription mistakes. We conducted *thematic analysis* as it allowed us to identify and interpret opinions, experience, and wishes that participants report. Using *ATLAS.ti*, we applied the six-step thematic analysis process by Braun and Clarke [79]. We based the initial codebook on the interview guide, memos, and research questions in a collaborative session. Using this codebook, two (of five) researchers initially coded each transcript independently. During coding, the researchers were free to add new codes. We reviewed those and merged codes or restructured the codebook if necessary in the joint discussion sessions, which we conducted in multiple rounds of 1–3 interviews each. In the next round, the coders continued with the updated codebook. Two researchers coded each interview transcript independently and later jointly discussed and resolved disagreements to a theoretical agreement of 100%. Therefore, we refrain from reporting inter-rater reliability (IRR) [80]. Overall, we assigned 4,488 codes, resulting in a median of 193 codes per interview transcript. We provide the final codebook as an online artifact (see Availability section).

### 4.6. Limitations

Our work has several limitations typical for interview studies and may suffer from biases, including over- and under-reporting, sample, and social-desirability bias. While we aimed to recruit a diverse sample of experienced and active researchers of the UPS community, our participants do not necessarily represent all UPS researchers, e.g., our

sample does not include researchers from the Asia-Pacific region. Us being part of the community and knowing many of its members may have introduced a selection bias. In addition, the researchers who participated may be more or less sensitized to transparency in research reporting or good scientific practices than those who declined. Some interview questions, e.g., about researchers' transparency practices, can be considered sensitive. Interviewing participants from one's own community and knowing some participants might also influence social desirability bias, similar to prior work [52]. To reduce social-desirability bias, we highlighted in the interviews that we were only interested in the participants' experiences and were not judging their answers. Overall, we perceived our interviewees as very frank and open (e.g., talking freely about poor transparency practices earlier in their career). Finally, we did not prompt participants for specific transparency practices. Therefore, and because of the qualitative nature of our results, the prevalence of practices may be a lower bound and may not generalize. Further quantitative studies need to investigate the prevalence of transparency practices in UPS papers.

#### 4.7. Positionality Statement

Following Ortloff *et al.*'s recommendation [77], we provide a positionality statement to be transparent about researcher characteristics, background, and position.

This project was a collaboration between two research groups in Germany and the U.S. The team comprises a full professor, an associate professor, a postdoctoral fellow, and two PhD students who are all UPS researchers. The latter three conducted and analyzed the interviews. The professors supervised the project. All researchers are experienced in conducting UPS research, and in writing, publishing, and reviewing UPS papers, and are familiar with typical UPS methods. The research team further included a research intern who holds a psychology bachelor and a high school intern. Both assisted with the qualitative data analysis. Overall, we believe that transparency is a crucial property of high-quality research and part of good scientific practice.

#### 4.8. Ethics

The human subjects review boards of our institutions approved this research. The research plan, study procedure, and all involved researchers adhered to German data protection laws, including the EU General Data Protection Regulation (GDPR). We followed the ethical principles of the *Menlo Report* [81]–[83] and the corresponding ACM policy [84]. We stored all data with PII in a secure, self-hosted storage and encrypted them at rest and in transit. Interviewees could familiarize themselves with the consent form and data handling information before the interview. We provided both on a website online. All interviewees consented to participate in the study, recordings of the interviews, and transcription of the recordings using a GDPR-compliant transcription service. Interviewees could skip questions or end the interview anytime. We conducted member checking

by sending preprints to all interviewees inviting them to suggest changes or correct quotes. Interviewees could contact us anytime for any questions or additional feedback and information.

## 5. Results

Below, we report on the six themes from the thematic analysis of our 24 interviews. We use quantifiers for the following percentile ranges to indicate prevalence of themes among participants: 0% = *none*; 1–20% = *a few*; 21–40% = *some*; 41–60% = *about half*; 61–80% = *most*; 81–99% = *almost all*; 100 = *all*. However, the analysis and findings are qualitative and should not be interpreted as representative quantitative results.

### 5.1. Varied Understanding of Transparency

At the beginning of each interview, we asked participants to elaborate on their interpretation of transparency in their work. While they reported no uniform, formal definition, they related similar aspects with transparency. However, we identified some differences in participants' reporting. A few participants mentioned that transparency is an ambiguous term without a clear definition:

*“It is a difficult question. I wouldn't say that I have an obvious and crisp definition. However, I can tell you the things that I think about when I think about transparency and the things that I personally try to do and encourage students and postdocs to do.” — P1*

A few others stated that there is no consensus in the UPS community on what is considered transparent: *“There's not a full consensus about what that entails, what's required, what the format of that is, and what it looks like.”* (P15). Participants also reported that the transparency of papers is not coherently assessed during peer review (Section 5.3).

**5.1.1. Purpose and Goal of Transparency.** More than half of the participants stated that transparency generally means to provide “all” or as much relevant information as possible in their paper, mainly to describe the research methodology and process. Most participants mentioned providing enough information and materials so others can attempt to replicate their work: *“For me, could someone read your paper and do the study that you did? I don't mean that in the sense of [reproduced] results, but could they do?”* (P19). Some participants aimed to make research verifiable, assessable, and generally understandable, e.g., for reviewers.

Participants also reported specific measures that make a paper transparent, like publishing research data, reporting IRR, discussing limitations, and further methodological details related to transparency. While participants mentioned various transparency practices, they predominantly reported sharing study protocols, questionnaires, interview guides, and codebooks (if applicable). Section 5.3.2 and Table 1 provide details on the interviewees' own and encouraged transparency practices.

In summary, participants mainly described the purpose of transparency from a methodological point of view—they intend to report all details and provide materials for others to understand, assess, and replicate the research fully. This is mostly in line with our definition from Section 2.1.

Our interviews mainly covered how to report UPS research transparently in a publication. They frequently mentioned providing data collection instruments, details on the methodology and research process, and sharing datasets if possible. Additionally, a few participants put transparency into a bigger context beyond individual papers, such as transparency towards study participants or in communicating science to the broader public. Some mentioned that publishing of replication studies and null results is a crucial transparency aspect (Section 5.5).

**5.1.2. Transparency and Related Concepts.** About half of the participants related transparency to other concepts, namely: reproducibility, replicability, and verifiability. A few particularly mentioned that transparency contributes to improved replicability and reproducibility (using both terms interchangeably), similar to our explanations in Section 2: “*I think reproducibility and replication you get as an added bonus.*” (P1). However, some explicitly stated that transparency alone does not imply reproducibility/replicability:

*“My research could be 100% transparent. [...] Someone could still have trouble reproducing what I did because they don’t have my experiences, research experience, research background, and the particular perspective that I bring. I’m not sure that’s a transparency issue.” — P17*

With that, the participant also highlighted the importance of researcher positionality. Other participants mentioned further limitations, e.g., having detailed descriptions of a study methodology but lacking access to the required participant population. Overall, our participants distinguished between transparency and reproducibility/replicability, and perceived transparency as a necessary prerequisite for both.

**Key Findings.** While exact understanding varied slightly, for many participants, transparency implied reporting all or as many details as possible, in particular regarding methodology. Additionally, most participants considered transparency essential for replicability and reproducibility.

## 5.2. Transparency is Good Scientific Practice, but Incentives Mismatch Drawbacks

Our participants reported several benefits of transparency. However, they also pointed out drawbacks and challenges hindering transparency. Participants agreed that a high standard of transparency is generally desirable, but most argued that investing the necessary time and effort is not sufficiently rewarded in the UPS community. Consequently, they felt one might be better off spending this time and effort on other projects: “*It’s easy to let [transparency] slip, there are no consequences for doing that. In fact, they’re probably better off.*” (P15).

**5.2.1. Motivations & Incentives are Often Grounded in Personal Beliefs or Benefits to the Community.** Some participants reported that transparency efforts depended on the motivation of individual researchers: “*It’s usually a matter of the capacity and interest of the individual PhD student working on the project.*” (P6). More than half of the participants reported a lack of incentives and motivation in the broader community: “*I think that that’s probably the biggest problem in terms of what’s blocking transparency: a general lack of incentives.*” (P15). A few participants mentioned that transparency did not provide actual benefits for individual paper authors: “*You don’t necessarily, as an author, get a ton of benefit from [sharing materials].*” (P18). Instead, half of the researchers indicated an intrinsic motivation to implement good scientific practices in their research and uphold scientific integrity as a reason to be transparent.

Other transparency benefits our participants mentioned pertained to the overall research field rather than individual publications or researchers. One such advantage was that transparency improves science, specifically by allowing for the assessment of the validity of research and finding potential flaws. Participants described how this potential scrutiny could improve scientific rigor and methodological execution during research:

*“If a project starts off thinking we are eventually going to make our raw data [...] [public], then it orients you toward transparency and good data organization, which leads to good, clean analysis.” — P23*

Consequently, a few participants argued that transparency increases the overall validity and credibility of research. Most interviewees also mentioned that transparency enabled easier replication and reuse of methods and data. Therefore, allowing the field to compare to and build upon prior work, and ultimately to progress faster.

A few participants pointed out how some of these advantages to the overall field translate to individual projects and papers. For example, a focus on publishing method details and materials can help structure a project and support good planning and execution, and a resulting transparent paper may be easier to publish. In addition, more than half of the participants stated that a paper whose method or artifacts get reused or built upon has a higher impact in the field. Yet, as one put it, “*that effect is very distant, somewhat disconnected, and very much not guaranteed.*” (P15). For example, even when transparency is sometimes recognized or rewarded by reviewers, this is not consistent and transparency could be a disadvantage, e.g., when disclosing imperfections or limitations results in negative reviews.

**5.2.2. Drawbacks and Challenges Often Affect Individual Researchers and Projects Directly.** Almost all participants cited the required time and effort, e.g., to publish and share materials, as a major barrier to transparency. One participant noted that deadlines and high workloads further hindered transparency efforts:

*“It’s also a lot of work, and given the fact that we’re a conference publication model and new stuff has to come out all the time and there’s a deadline, sometimes it’s hard to get what seems like extra work done.” — P7*



Almost all participants felt that being transparent with methodological details made them vulnerable or put them at a disadvantage in the peer-review process. Some pointed out how flaws or minor mistakes might become more apparent, and sometimes open them up to unwarranted reviewer criticism. For instance, details in the study materials might capture reviewers' attention and could disproportionately diminish the paper's perceived validity or contribution. One participant summarized the effects:

*"Everybody is in such a defensive mode, like do not give people an attack surface. When you write a paper, don't give reviewers this attack surface. If we don't change this, we will never ever be able to get full transparency."*  
— P2

Both spending extra time and effort on transparency and exposing oneself to increased criticism from reviewers put the publication performance of researchers at risk. However, they perceived publication rate as one of the most important success metrics in academia in general and in UPS, which decides job market prospects as well as third-party funding. Correspondingly, many participants regarded this as a severe disadvantage to being transparent. Therefore, participants thought it was unlikely that researchers would go beyond current community practices without stronger incentives for being more transparent.

Furthermore, participants reported challenges regarding publishing research materials and data. Almost all researchers discussed ethical considerations, data protection, and the protection of company-owned data, or intellectual property (IP) they worked with as restrictions for publishing data. Specifically, they worried about accidentally disclosing personal information of study participants, especially in qualitative data and when working with vulnerable populations. About half cited difficulties with de-identification techniques and assessing their level of protection: *"I never feel confident enough that we will have reduced identification risk sufficiently."* (P6). As a result, participants did not publish data sets that contained or might reveal personal information of participants.

All participants had encountered constraints to transparency, the most common one being page limits for papers and appendices. While all participants agreed that page limits could harm transparency, about half of them said they were necessary to prevent verbose publications and maintain readability. Some argued that page limits did not actually prevent transparency because external supplementary materials were an option, and thus removing them would not automatically improve transparency: *"For me, it always was, if you want to publish transparent research that won't fit in a paper, there is some additional documentation that you need to provide."* (P9). Another constraint was the need to stay anonymous during the reviewing process. Over half of the participants described technical difficulties in providing materials. These included ensuring the long-term availability of materials and handling large data sets and code that require extensive resources to run. A few participants desired standardized solutions and tools to problems like hosting archives, preparing materials, and de-identifying data.

**Key Findings.** Overall, we find a mismatch of incentives for transparency and its drawbacks and challenges that need to be overcome to enable better transparency in UPS research. Besides better incentives for transparency, we find a need for solutions to help individual researchers overcome transparency challenges.

### 5.3. Overall State of Transparency Has Improved, but Further Progress is Desired

Almost all participants reported observing an improvement of transparency in the UPS community over time. However, most mentioned that further improvements are desired to overcome existing obstacles. For instance, one participant reported on the current transparency state:

*"Moderate, but better than when I started, but still lots of room for improvement. One example of what I think is improving is that it is now fairly common practice [...] to publish the questionnaires in their appendices. That was the exception when it started, and now it's almost becoming the norm. I've seen reviewers penalize papers submitted to SOUPS for not having the study protocol in appendices, and I know that I do that."* — P4

**5.3.1. Remaining Problems with Transparency.** The desire for further improvement is likely grounded in several problems with transparency that the participants experienced. Most participants criticized missing methodological details in many papers, for example: *"That was eye-opening. We read just method section after method section [...] [and saw] across the board how little information is sometimes included."* (P19). As participants reported, this includes missing participant recruitment details, a lack of study materials (e.g., questionnaires, interview guides), and underreporting of critical procedural steps. The unavailability of details and materials also is a major challenge when replicating published papers, as some participants reported: *"Occasionally you'll want to replicate something old where the actual survey questions, or whatever, were not included. [...] If that information is not included, it can be a challenge."* (P11). And availability is not enough, as a few participants also experienced available but unusable materials, e.g., program code that did not run.

Half of the participants reported that reaching out to paper authors is a common practice to request otherwise unavailable information or materials. However, this can provide challenges for researchers:

*"Over the years, there have been studies I wanted to replicate. Often that information isn't available. You try emailing the original authors, and they're, 'Hey, let me find my backup drive and see if I can find that for you. It's long since been deleted.' They can't, right?"* — P5

About half of them reported further challenges, such as authors not archiving, losing, or not finding data, or authors not being reachable anymore due to leaving academia or losing access to old email addresses. Contrary, some participants reported positive experiences such as *"Overall, if you email the authors [...], they can generally find those things."* (P12) or *"within a couple of hours they sent me the data."* (P4).



**5.3.2. UPS Community's Transparency Practices.** Participants desired transparency in UPS, and tried to counteract the problems described above with transparency practices. Those included both general principles participants follow and concrete steps they implement. General principles include reporting research questions, justifying methodological decisions, providing methods in detail, disclosing study limitations, and publishing data (if possible)—or generally “*striving to essentially provide as much of this information as possible just in case someone asks for it.*” (P24). While the above are generally good scientific practices, participants reported specific practices for UPS research that depend on the type of study.

Table 1 summarizes participants' transparency practices, differentiating between those practices participants stated to follow themselves (*own practices*) and those they generally encouraged (*encouraged practices*). These are not mutually exclusive, i.e., many participants stated to both implement practices in their research and generally encouraging them (e.g., as a reviewer or in their research groups). Overall, there is no major difference between participants' own and encouraged practices. In fact, participants' own practices are largely a subset of encouraged practices, as participants naturally encourage what they implement. As we did not prompt all practices, participants might underreport common or obvious ones, not explicitly mentioning a particular practice that they implement. For example, participants rarely mentioned reporting institutional review board (IRB) approval, while UPS papers typically contain such sections and this is often a CfP requirement.

Frequently mentioned transparency practices concerned describing recruitment and other aspects of human subject studies, including ethical considerations. For example, some participants noted the importance of describing recruitment and demographics:

*“The recruitment methods, who is in the study, all of those things are so important to have in the paper so that people understand the context for the insights that come out of the research.” — P19*

However, participants rarely mentioned to report some recruitment details such as compensation or inclusion and exclusion criteria. Besides reporting research methods and results, participants also provided research artifacts and encouraging it. Depending on the study type, participants most frequently reported providing codebooks for qualitative studies, interview guides, online questionnaires, analysis scripts, or similar materials alongside their papers. Related, participants reported to regularly provide datasets if possible, or if having permission by participants. Most participants also provided additional information and data upon request, besides what they include in appendices and online supplementary materials. Besides these openness and availability aspects, participants also reported some transparency practices for research integrity, such as preregistration, disclosing imperfections or any method changes, publishing null results, and archiving data.

**5.3.3. Transparency in Peer-Reviewing.** The peer-review process critically impacts transparency practices in published papers. According to our participants, transparency is often considered but not always a major concern in reviewing and is highly dependent on specific reviewers. Overall, half of them experienced little to no consensus among reviewers regarding transparency, matching the varied understanding reported above (Section 5.1): “*No, I don't think there's a consensus. As the quality of reviews varies, the notion of transparency varies as well.*” (P14). Participants reported that non-UPS reviewers and differing understanding between sub-fields might cause disagreements.

Some participants reported a positive effect of the reviewing process on transparency:

*“leaving out the actual [...] questions, leaving out details. This is something that I see a lot more in papers I'm reviewing than in the actual published papers. A lot of this gets corrected in the process” — P5*

Fittingly, most participants expressed that transparency considerations are essential in their reviewing activities: “*It's definitely a major consideration. It sets the whole tone for whether or not I believe their results.*” (P14). A few stated the same on reviews they received: “*I've always felt that it's important for the reviewers.*” (P21). Some participants explicitly stated that only transparent research reporting allows reviewers to assess the validity of reported methods and additional materials: “*I expect to see the survey instrument. [If it's missing,] that's a big red flag because then I don't really know how to evaluate the paper.*” (P21).

In that sense, some participants mentioned that transparency makes the evaluation easier and improves review quality. Lacking transparency contributes to lower quality reviews: “*I need this information in the paper. I can't evaluate it because I'm not going to evaluate it based on being missing in the main paper.*” (P11). Consequently, more than half of the participants indicated that they asked for missing information or materials in their reviews and the rebuttal process. However, some stated that transparency is one of multiple aspects in their review or only mentioned when lacking, and, therefore, not of primary concern: “*For me, transparency plays a role but there are so many other factors. It's a bit hard to position it and it really depends on the individual case.*” (P2). On the contrary, a few other participants make transparency a top priority in their reviews and try to reward authors for good transparency: “*I'm impressed by that. You get me fighting just a little bit harder for you.*” (P16). Considering all these insights, paper authors might get better reviews and avoid negative feedback when being transparent.

**Key Findings.** Participants reported an improvement of transparency in UPS, but desire further progress. Transparency issues include lacking materials and method details, which limit replicability and prevent reviewers from accurately assessing papers. While bad transparency might result in negative reviews, researchers fear providing an attack surface for reviewers when being transparent.

TABLE 1. PARTICIPANTS' OWN AND ENCOURAGED TRANSPARENCY PRACTICES (AS REPORTED WITHOUT PROMPTING EACH INDIVIDUALLY).

Transparency Practice	Description	Own*	Encouraged*
<b>Recruitment &amp; Participants</b>			
Describe demographics/sample	Include any information that describes the sample (e.g., gender, age, education, race, ethnicity).	●○○	●●○
Describe recruitment	State how recruitment/sampling was done, including how participants were contacted. If applicable, provide recruitment materials (e.g., emails).	●○○	●○○
Report inclusion/exclusion criteria	State any inclusion/exclusion criteria and report the number of excluded/included participants.	○○○	●○○
Report compensation	Describe if, how and how much the participants were compensated for participating.	○○○	●○○
<b>Qualitative Research</b>			
Provide interview guide	For interviews and similar studies, report the original interview questions and structure.	●●○	●●○
Provide codebook	For qualitative analyses, provide the codebook that was used. Additionally, state the source of the codes or how they developed, and some example quotes.	●●○	●●○
Describe qualitative analysis in-depth	For qualitative research, describe the codebook creation and qualitative coding process in detail.	●○○	●○○
Provide positionality statement	Especially for qualitative work, explain the researcher positionality regarding the study topics.	●○○	●○○
(Don't) Provide interview transcripts	Providing interview transcripts can be debated. While desirable, even when anonymized, it might still be possible to identify participants based on the larger context.	●○○	●○○
State inter-rater reliability (IRR)	For qualitative research (if applicable), discuss IRR and how it was measured.	○○○	●○○
Include interview quotes	For qualitative work like interviews, provide (anonymized) quotes as evidence to support findings.	●○○	●○○
<b>Quantitative Research</b>			
Describe statistics	For quantitative research, explain the statistical analysis and method in detail.	●○○	●○○
Provide analysis script	Provide any scripts that were used for data analysis (e.g., <i>Jupyter Notebook</i> ).	●○○	●○○
Provide all statistical results	For quantitative research, make sure to report all results (significant as well as insignificant) and statistical key figures (e.g., confidence intervals, effect sizes).	●○○	●○○
Make analysis traceable	Each analysis step should be traceable (e.g., Python script). Ideally, there is no manual step between data input and computed results (e.g., statistics, figures).	●○○	●○○
<b>Writing &amp; Paper Content</b>			
Clearly state research method	Include all information on the research methods and the overall process that was followed.	●●○	●●○
Write justifications	Explain the why behind decisions, e.g., why the specific method was used or data cannot be published.	●○○	●○○
Provide clear definition of terms	Avoid ambiguity by defining terms (for both readers and participants).	○○○	●○○
Clearly state research questions/goals	Aim to clearly describe a study's RQs or alternatively a research goal.	●○○	●○○
Describe piloting	State whether and how the study was piloted, including any changes made based on piloting.	○○○	●○○
Discuss limitations	Name and explain any study limitations and if/how they were mitigated.	●○○	●○○
Get feedback on paper transparency	Ask others to review paper for transparency, e.g., check for missing materials or information.	●○○	○○○
<b>Research Integrity</b>			
Study preregistration	Register the research question and planned methodology publicly before collecting data.	●○○	●●○
Take notes & document research process	Keep track of the original research idea and its development during the research process, e.g., by writing down the individual steps, decisions and their rationales, and any method changes.	●○○	●○○
Disclose imperfections/problems/changes	If some issues arose, or even changes to the research method were necessary, those should be disclosed.	●○○	●○○
Don't change research questions	RQs should not be changed, e.g., to prevent fitting RQs to results post-hoc.	●○○	●○○
Publish null results	Include any negative results from a study to prevent positive result bias.	●○○	●○○
Use registered reports	If a venue adopted <i>registered reports</i> , consider submitting a preregistered study proposal.	○○○	●○○
Store data and information persistently	Archive any data and all information regarding the research project.	●○○	○○○
<b>Data Collection</b>			
Provide data collection instrument	If applicable, provide any instrument or protocol that was used for data collection.	●○○	●○○
Provide questionnaire	For surveys, provide the questionnaire with the exact original questions.	●●●	●●●
Publish data collection code	Publish the program that was used for data collection (e.g., crawler). Especially in cases where publishing original data is not possible, this can improve transparency.	●○○	●○○
Provide code/software publicly	Provide any used software (e.g., tested prototype) and its code open source. If not possible, consider screenshots or a video.	●○○	●●○
Describe study context	State in which context a study was done (e.g., the environment, duration, location, date).	●○○	●○○
Report data cleaning	For example, describe when data is considered invalid or incomplete.	○○○	●○○
<b>Research Data Sharing</b>			
Provide research data	Share the original (anonymized) study dataset if possible.	●●○	●○○
Obtain consent to release data	Ask participants for consent to release (de-identified) data, e.g., to publish sensitive data.	●○○	●○○
Publish synthetic data (instead of PII)	When providing sensitive data/PII is not possible, one could provide synthetic example data.	●○○	●○○
<b>Openness &amp; Availability</b>			
Provide appendices/supp. materials	Include materials used in the study (e.g., questionnaire) and artifacts created in the study (e.g., analysis script) either in an appendix or online.	●●○	●●○
Provide materials/information on request	Respond to questions (e.g., via email) of other researchers needing information or requesting materials/data.	●●○	●○○
Make research freely available	Publish papers open access if possible or make them available on the authors' personal/institutional websites.	●○○	●○○
Link online materials from paper	Ensure to include a (clickable) link for easy access to any online materials.	●○○	●○○
<b>Ethics &amp; IRB</b>			
Disclose Ethics Approval (e.g., IRB)	Disclosed whether ethics approval was obtained and what the result was.	●○○	●○○
Discuss ethics	State any ethical considerations and limitations.	○○○	●○○
Provide IRB/ERB materials	Include documents, for ethics application and approval.	●○○	●○○
Report vulnerability disclosure process	If reporting vulnerabilities, describe the (responsible) disclosure procedure. This could include original materials (e.g., emails).	○○○	●○○
Provide consent form	The form or any information presented to participants to obtain informed consent.	●○○	●○○
<b>Miscellaneous</b>			
Establish transparency within research team	Communicate transparency practices clearly and promote them inside the research team.	●○○	●○○

\*Prevalence of practices in our interviews (in number of participants): none (0) = ○○○; low (1–8) = ●○○; medium (9–16) = ●●○; and high prevalence (16–24) = ●●●.

## 5.4. Transparency is Directed by Implicit Standards, Lacking Formal Guidance

Most participants reported lacking guidelines or official requirements for transparency: “*I don’t think there’s any written requirement. [...] I think it’s more of a cultural standard or a norm.*” (P12). While a few mentioned some guidelines, the 2010 guidance on how to write a SOUPS paper [9] was named only once and is the only one specific for UPS. The remaining participants reported internal, self-written publication checklists and a guideline on statistical reporting that a venue from another field suggests [54]. Overall, most participants largely observed an implicit, not formalized, transparency standard in the community:

*“Requirement is an interesting word because I think they’re de facto requirements. It’s maybe not written down necessarily in the CfP all the time, but it’s sort of a requirement. The reviewers will expect it.”* — P11

### 5.4.1. Formal Transparency Guidance would be Helpful.

Most participants generally liked the idea of having written, formal transparency guidance, e.g., alongside or within CfPs. Some mentioned that clearly communicated guidance makes venues more inclusive for new authors who might not know implicit community norms:

*“[I]t’s important to try to write some of the stuff down. [...] it would be really nice and important to provide some more guidance for new people and not make it so that if you’re not already in the club, you don’t know the rules, and now you can’t publish the research.”* — P11

Others mentioned that written guidance helps reach a community consensus, makes transparency more actionable, and can set transparency incentives. A few participants suggested that uniform requirements improve the fairness among researchers by requiring the same standards, as in the following example: “*I do think we might eventually need to have it as a requirement just to equalize.*” (P6).

**5.4.2. Requirements vs. Recommendations.** How to give guidance—as mandatory requirements or optional recommendations—is debatable. Some participants had concerns about adding formal requirements to CfPs, mainly because enforcing requirements might have downsides. Overall, about half of the participants discouraged hard requirements because some situations might need exceptions:

*“There should be guidelines and recommendations. I don’t think there should be requirements necessarily because there are always exceptions. So I like it to be a guideline and recommendation.”* — P10

Examples of such a situation might be sub-disciplines of (U)PS, an unconventional study type that does not fit the transparency guidelines, or a study that does not disclose detailed demographics to prevent identification of (at-risk) participants. However, most participants mentioned that at least some guidance could be hard requirements: “*In some cases, I think it’s okay to make a requirement. I feel like it’s okay to demand survey materials.*” (P21). Actually, participants mentioned examples of hard requirements that

improve transparency, e.g., CfPs that require ethics statements and funding organizations that require researchers to provide results and data publicly.

In line with the preference for recommendations instead of requirements, some participants asked to introduce transparency best practices as optional recommendations gradually. Supporting this, some participants stated that changing community norms is difficult and takes time anyway. With a gradual introduction, they argued that recommendations can still become requirements later when the community discusses and agrees on those and do not harm any researcher when suddenly enforced. These social side effects should be considered to not discourage researchers to be transparent. One said:

*“I think what’s more important is the community deciding what these norms should be and why and then making sure that people are following that and encouraging without necessarily browbeating people.”* — P18

**5.4.3. Transparency Guidance for Reviewers.** As providing materials and information is part of transparency, we asked participants whether reviewers should be required to consider appendices or any other (online) supplementary material during the review. Currently, at most venues, reviewers are not required to consider appendices. Most participants agreed to keep this policy for various reasons, including reviewer load, reviewers not being experts in artifact review, papers’ main parts being self-contained, and that nobody could check whether reviewers actually look at appendices. One participant proposed changing the policy to account for the authors’ additional effort: “*I think if we ask authors to produce this stuff, then we should ask reviewers to review this stuff properly.*” (P1). Other participants proposed that reviewers should be encouraged to consider transparency and be provided with guidelines, that there should be a dedicated appendix reviewing committee, or that a single reviewer could be designated to check a paper’s appendices and supplementary materials. Some participants suggested introducing a field or checkbox in review forms (e.g., in HotCRP) for transparency or material availability (analogous to how ethics are handled)—either in detail or just to check availability:

*“At the checkbox level, is a survey instrument included in the appendix? That could just take a 30-second check. You don’t have to necessarily review that instrument and read it line for line. Some might. I don’t think reviewers should be expected to read the appendix.”* — P21

**5.4.4. Wishes & Recommendations for Future Transparency in UPS.** Participants made recommendations and expressed wishes towards improved transparency within UPS. This included encouraging practices, such as those listed in Table 1, and largely overlaps with the practices that participants reported following themselves (Section 5.3.2). In that context, the main wish was guidance on transparency, as participants agreed that the UPS community would benefit from best practices and clear expectations (Section 5.4.1).

Several wishes concerned the UPS community culture. While some mentioned that venues should shape trans-

parency, others stated that the whole community has to build a transparency culture, starting in individual research groups, and teaching students transparency. Our participants felt that transparency should be appreciated more in reviews and called for transparency, replication, and reporting null results in CfPs.

Finally, more than half of the participants noted the need for more transparency incentives, to overcome the barrier of added effort that transparency imposes: *“I think mostly incentives. [...] that’s how academia works, for sure. If you create incentives for people to do something, then for sure they are going to do it.”* (P24). Concrete incentive proposals were awards for transparency, special venues or publication options for transparency-related artifacts, and considering transparency in academic hiring.

**Key Findings.** We found that implicit community norms impact transparency in UPS, but formal guidance is missing. Participants wish to have written guidance as optional recommendations, not hard requirements. They suggest that reviewers focus more on transparency, while not having to review appendices. Other wishes included concrete incentives and a mindset shift in the UPS community towards promoting transparency.

## 5.5. Null Results and Replication Studies are Essential but Hard to Publish

Beyond transparency of individual papers, participants highlighted that publishing replication studies and null results contributes to transparency overall.

Our interviewees had mixed experiences with replication studies. While some reported never having tried to replicate previous work, a few successfully published replication studies, and some attempted replication without submitting their results or not getting the paper published. Nevertheless, half our participants emphasized the significance of replications: *“I believe vibrantly and deeply in the need for replication and reproduction of research”* (P16). Simultaneously, more than half perceived a general lack of replication studies in the UPS community. In contrast, a few participants stated that replication is not universally useful overall, if not providing additional insights. A few participants reported general difficulties of publishing replications: *“often you replicate something, and you find the same results. In which case it’s hard to get the reviewers excited about it.”* (P4).

Others emphasized the significance of null results: *“It’s so important to get that out there because otherwise, other people will try that and also get null results.”* (P16). Some participants highlighted that publishing null results is necessary for transparency. As one put it: *“If we want to have the overall picture and not just the tiny snapshot of the study that showed some positive results, we need to be transparent about negative results as well.”* (P13). Interestingly, only one participant questioned whether null results are always useful, e.g., when null results are expected and therefore less relevant.

Some researchers reported it to be challenging, if not nearly impossible, to publish null results. They commonly

perceived the UPS community to heavily favor positive findings, possibly impacting the findings:

*“If you found a null result, the temptation to reinterpret that to produce a non-null result is very, very high. You will get your paper published. It’s what we train researchers to do, as the goal is problematic.”* — P1

While, some participants identified lacking incentives for both replication studies and null results as factors that reinforce the issue, one participant disagreed with a publication bias and argued that the UPS community is different from other fields in this aspect. More than half of our participants felt that preregistration and registered reports might enable null result publication: *“If you end up with negative results or you end up with no results, [...] you can still publish it because you can say this was approved”* (P18).

**Key Findings.** While not all participants mentioned the topic, the majority concurs that conducting and publishing replication studies and reporting null results are important for transparency. However, many felt that UPS lacks both due to a publication bias that provides a disincentive.

## 5.6. Artifact Evaluation is Promising, but Needs to be Adapted to UPS

Almost all participants were familiar with AE, but only some had actively participated. Only a few expressed insights from other communities, yet some participants perceived AE as an emerging trend in the SP community: *“The security communities are starting to have [...] these artifact or artifact evaluation programs. It is a shift towards transparency”* (P12).

### 5.6.1. Current AE is not Directly Applicable in UPS.

More than half of the participants generally appreciate AE, and some think that it improves reproducibility. They reported several positive experiences of AE such as *“a better chance at having reproducibility, particularly when it comes to code and verifiability when it comes to data sets”* (P16). Despite liking the idea of AE, about half also expressed concerns. Specifically, most concerns revolve around the suitability of AE for UPS research. When implementing AE in UPS research, participants found a lack of clear expectations that results in hesitancy to participate in AE. As one of the participants puts it: *“I have always been mildly confused about what an artifact is and whether or not [...] our research outputs qualify as artifacts”* (P6). While some SP venues have AE, participants found that it is mainly tailored towards software artifacts, which often does not apply to UPS artifacts, such as survey instruments. These statements fit the vague AE policies for artifacts and replicability of human subject studies (cf. Section 2.2). Our participants expressed wishes that it should be specified what artifacts are expected for a certain study type. Other AE concerns extended to data protection when sharing data, which might not be as critical in other SP studies that involve no participant data: *“For any interview data, I would be very concerned about like how we’re transferring the data.”* (P6).

### 5.6.2. AE is Labor-intensive and Should be Rewarded.

Another issue according to about half of our participants is the extra work that AE causes. This applies to both authors, who have to prepare the artifacts, and the evaluators who review them. The latter might be hard due to the normal reviewing load, as a participant stated: *“I wish we would have more artifact evaluation going on, but honestly, as a reviewer, I’m swamped with reviewing requests all the time.”* (P2). Therefore, a few participants requested that the additional efforts of AE have to be *“suitably compensated or rewarded”* (P5) for both, authors and evaluators. Some recognized the AE badges as a reward: *“you should be able to put on your resume that you were doing artifact reviews and stuff. And that should be something that’s viewed as community service for tenure reviews”* (P5). One participant half-jokingly proposed to pay evaluators to implement AE widely, like some journals pay their reviewers: *“You’d have to pay people to get them to do all that work.”* (P7). Due to the difficulties mentioned, our participants did not desire to make AE mandatory. Additionally, some participants emphasized the need in UPS research to provide artifacts alongside the paper during the review process rather than in a separate AE process: *“As a reviewer, I really want to see some of these transparency artifacts”* (P15).

**Key Findings.** Participants perceived AE as promising to improve transparency. However, they struggle with the applicability of current AE at SP venues to UPS research. The participants emphasized the need for clearer expectations and to adapt AE to non-technical artifacts, which are typical for UPS research. The additional effort was a concern that needs adequate rewards for authors and evaluators.

## 6. Discussion

We discuss the key takeaways and implications of our findings and provide recommendations for further improving transparency in UPS research.

**Novelty & Significance of our Findings.** In long-established fields such as medicine or psychology, researchers have investigated research transparency and developed several frameworks and principles for it (Section 3.2), e.g., establishing study preregistration [85], [86] in response to the declared reproducibility crisis [23], [24]. These efforts aim to strengthen the rigor and reliability of scientific research and its findings. The broader SP community has started adopting AE for research transparency, and the AI community started utilizing mandatory transparency and reproducibility checklists in response to identified transparency problems [8], [32]–[34], [87]. The UPS community also still continues to develop its research practices and culture, but meta-research into those practices only began to emerge recently [44], [45]. So far, it has been unclear to which extent findings and recommendations from other fields have found application in UPS research. As a core principle of good science, the (U)PS community needs to consider transparency as part of general meta-research to advance its scientific pursuit [1], [2]. With our researcher

interviews, we contribute first qualitative insights on UPS scholars’ perspectives on transparency. Based on these perspectives, we make specific recommendations to overcome challenges, foster discussion towards community consensus on best practices that align with the community culture, and incentivize transparent research reporting in UPS.

**Transparency Practices (RQ1).** We find, and thereby can further support, practices stated by our participants, that are similar to the recommendations in other sub-fields of computer science, such as archiving data and information, rewards for providing artifacts [33] and documenting the research process [8]. At the same time, we find other important practices that emerge from researching humans and using mixed methods that are not present in other CS subdisciplines. While some transparency practices we identified align with those in other fields, we find that not all of them are widely embraced by the UPS community yet, according to our participants. For example, UPS researchers commonly describe their recruitment approaches as suggested already in 2010 by Schechter [9]. However, there appears to be less consensus and a lack of written guidance on how to report qualitative UPS research, even though other fields have developed respective reporting expectations [59], [88]–[90], or for obtaining consent for data publication [91]. Therefore, our findings provide first insights on the current state and the community perspective of transparency in UPS. Overall, our participants reported various transparency practices (Table 1) such as providing study materials (e.g., questionnaires, interview guides), statistic tables, and software artifacts. However, those were predominantly informed by perceived community expectations rather than formal guidelines (Section 5.4), which can lead to inconsistent reviewer feedback and pose entry barriers for researchers new to the field. We argue that the UPS community should work toward establishing explicit transparency guidelines to avoid arbitrary variability in reporting and reviewing. Our insights can inform data-driven development of future transparency measures and guidelines.

**Transparency Benefits & Challenges (RQ2).** Many participants perceived an imbalance between the incentives and drawbacks of transparency (Section 5.2). Our findings for UPS align with lacking transparency incentives reported in other research fields [5], [50]. However, there are concrete adjustments our community could make to address this imbalance. For example, we find that current AE processes at SP venues might prevent authors from participating if they are not well aligned with artifacts produced in UPS research (Section 5.6).

**Transparency Development (RQ3).** While participants attested to an improvement in the UPS community’s transparent reporting in the past, they also highlighted ongoing transparency issues (e.g., missing information and materials). They identified multiple areas for future improvement (Section 5.3). Some of our findings align with meta-research on transparency and replicability in other communities [36], [42], [43], demonstrating the relevance of meta-research insights from other fields for UPS, while others are UPS

specific. Our findings further confirm certain insights from less-comprehensive related work, e.g., on risk representation and participant demographics in UPS research, which highlighted poor research reporting [44], [45]. McDonald *et al.*'s influential HCI paper on IRR showcases the impact of a single paper on community standards and reporting practices [80]. We hope to initiate a similar discussion towards improving transparency in the UPS community, replacing implicit expectations with clear guidelines.

## 6.1. Recommendations

Based on our findings, we provide recommendations for improving UPS research transparency through community efforts.

**6.1.1. Developing Transparency Guidance for UPS.** Our participants desired formal transparency guidance for UPS publications, pointing out its potential to cultivate a community consensus and increase fairness and inclusiveness in the UPS community by reducing dependence on institutional knowledge and resources (Section 5.4). Prior efforts, such as Schechter's [9], have not been maintained and are outdated. A community effort is needed to develop a contemporary set of core transparency requirements for UPS publications. Our interviews suggest that these core requirements should result from a discussion in the UPS community, e.g., at a workshop, to ensure applicability in practice and improve acceptance. As a starting point for this discourse, we contribute the overview of transparency practices encouraged by our participants in Table 1. Commonly already submitted and expected materials such as interview guides, survey questionnaires, and experiment materials are good candidates for more formal requirements. In addition, low-effort practices, e.g., providing recruitment materials, should be considered. We further recommend extending these requirements with recommendations for additional transparency practices. To help build a consensus and foster community awareness, we propose collecting requirements and recommendations in an independent project with a website for venue CfPs and reviewing instructions to link to.

Publication venues could then adopt and distribute formalized guidance on transparency, e.g., by adding them to CfPs or reviewing instructions. Whether this guidance should take the form of requirements or recommendations remains a debated topic (Section 5.4.2). Requirements could incentivize compliance and reliably raise the general standard of transparency. However, our participants expressed concerns that requirements may be inflexible and could result in disproportionate punishments in edge cases that were not considered during the requirements' development, e.g., exceptional methodologies or circumstances. For example, some SP venues introduced mandatory ethics considerations in response to the 2021 *Hypocrite Commits* incident [92]. Unfortunately, ethics requirements and researchers' reality do not always match [93].

**6.1.2. Incentivizing Transparency.** In addition to the community establishing transparency standards, we also see exciting opportunities for publication venues to better incentivize authors to make their publications more transparent (Section 5.2).

**Artifact Evaluation.** First, our participants identified AE as a promising avenue that needs improvements to be useful for the UPS community (Section 5.6). We suggest that venues continue and extend their efforts to support AE, and adapt it to the needs of UPS (and other subfields) by refining the description of accepted or expected artifacts for different types of studies. Participants recognized AE badges as rewards. We propose introducing new badges for different study types, e.g., "study materials provided."

**Transparency Awards.** In 2013, ACM CHI introduced the *RepliCHI Award* for publications with high-quality replications or high replicability [94]. While unfortunately discontinued, we encourage venues to review lessons learned and consider introducing similar awards to incentivize and reward authors of especially transparent publications. Listing transparency awards on CVs could distinguish researchers, and might be recognized in academic hiring and promotion.

**6.1.3. Addressing Challenges.** Besides providing incentives, publication venues could also better support authors in their transparency efforts by addressing challenges that our participants reported (Section 5.2).

**Removing Appendix Page Limits.** Participants expressed that page limit constraints frequently hinder transparent reporting (Section 5.2.2). With digital publishing [95], appendix page limits should be lifted, like PETS/PoPETs did [96], or replaced with sufficiently generous limits. This provides space for publishing study materials as part of a paper, as desired by many participants—while preserving limits on the paper's main body that promote brevity.

**Providing Artifact Hosting.** Venues or research organizations (e.g., IEEE or USENIX) could provide reliable long-term hosting platforms for research artifacts, or promote and support projects like *Security Research Artifacts* [73] or Open Science Framework (OSF). Artifact hosting platforms with long-term archiving can ensure availability for decades—which is unlikely for personal, self-hosted websites. The ACM already offers hosting supplementary materials alongside papers in the *ACM Digital Library*.

**Alleviating Ethical, Privacy, & Legal Concerns.** Data-sharing can be especially hindered due to ethics, privacy, and legal constraints. Typically, UPS research involves participant data. In line with related work [53], interviewees outlined that they avoid sharing sensitive participant data due to ethical and privacy concerns. Other legal aspects can also prevent sharing, e.g., industry collaborations that prevent publishing data due to IP restrictions (Section 5.2). The field of social sciences has developed solutions for data archives containing PII and corresponding access and privacy management [97]. Similar avenues might be worth exploring for UPS and promoting in the community.

**Reducing Effort Through Tooling.** We argue that the UPS community needs supportive tools to reduce transparency efforts. For example, it currently requires multiple manual steps to include an online survey, e.g., from Qualtrics, in an appendix. Instead, a script to convert the survey to  $\text{\LaTeX}$  would be quicker and easier. Respective tools are currently missing, but may be meaningful community efforts.

**6.1.4. Reviewing & Submission Platforms.** Publication venues would potentially need to rely on the reviewers of a steadily rising number of submissions [52] to enforce and implement emerging transparency guidelines and requirements during the review process. Thus, it is important to carefully consider how each measure impacts the (already high) reviewer workload and offer respective support where possible. A feasible initial approach might be to introduce a transparency checkbox or field to review forms (e.g., in HotCRP) so that reviewers can easily indicate whether a submission is sufficiently transparent and provides all relevant information and materials. Similarly, review forms could offer an option to nominate submissions for a transparency award. Regarding artifact evaluation and hosting, we suggest integrating a field for anonymized links to external repositories, e.g., as offered by OSF or providing upload fields to submit artifacts alongside the papers. This helps to ensure that the review is double-blind and artifacts are available at the time of paper review—making the process easier for both authors and reviewers. However, only a few venues currently do this, e.g., ACM CCS or ACM CHI.

## 7. Conclusion

We qualitatively investigated transparency practices in the usable privacy and security (UPS) community based on 24 interviews with experienced UPS researchers. Although the UPS researchers value transparency overall, a lack of incentives, additional effort, and potential disadvantages, like offering an attack surface for reviewers, hinder transparency. While participants reported that the community follows an implicit set of community expectations around transparency, that has led to transparency improvements over time, they desired further progress as part of good scientific practice to improve UPS research quality. The UPS community needs to engage in a constructive discussion, which we hope to start with this paper, (i) to develop guidelines that set explicit standards for transparency and (ii) on how to set appropriate incentives. We argue that some lightweight measures can significantly improve transparency at little cost, such as removing appendix page limits and including study materials in the appendix. While we focused on UPS, our insights can be useful for the broader SP community, and might (partially) apply to other fields, such as HCI.

*“It is more effort, but what transparency is intended to do is produce more rigorous research, and I think if we don’t produce rigorous research, then what’s the point? It is just a foundation for what we do as researchers.”*

— P17

## Acknowledgments

Most importantly, we thank all fellow UPS researcher who participated in our interviews for their time—despite busy schedules—and serving our community with their insights. With this research, we want to give back to the UPS community by improving transparency and related aspects such as replicability and reproducibility. We thank the anonymous reviewers for their valuable feedback and for helping us to improve this paper. The Hannover authors want to thank our feel-good manager Cleo for being the cutest office dog and always offering emotional support. This research was partially funded by VolkswagenStiftung Niedersächsisches Vorab – ZN3695 and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy – EXC 2092 CASA – 390781972. Byron M. Lowens has been supported by a CRA/NSF CI Fellowship under grant No. 2030859.

## Conflicts of Interest

The authors declare none.

## Availability

In writing a paper on transparency, we aim to lead by example. Therefore, we included as many details and examples as possible. For the reporting, we followed the *SRQR* [59], *COREQ* [88], and the relevant *Empirical Standards* [70] guidelines. However, the paper length is limited, and we could not fit all of it in this paper. Therefore, we provide artifacts in the appendix and online. While desirable for transparency, we refrain from providing the interview transcripts and their coding because even anonymization might provide enough context to identify participants. Our replication package contains the following artifacts: (1) The complete interview guide, (2) the demographic questionnaire, (3) the recruitment materials, (4) the project website, including our consent form, (5) the codebook, and (6) the dataset of UPS paper used for recruiting. The replication package is available at: <https://doi.org/10.17605/OSF.IO/HY7EJ>

## References

- [1] R. Schell, “Information security: science, pseudoscience, and flying pigs,” in *Seventeenth Annual Computer Security Applications Conference*, 2001, pp. 205–216.
- [2] 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.
- [3] M. Baker, “1,500 scientists lift the lid on reproducibility,” *Nature*, vol. 533, 2016.
- [4] National Academies of Sciences, Engineering, and Medicine, *Reproducibility and Replicability in Science*. The National Academies Press, 2019.
- [5] M. R. Munafò, B. A. Nosek, D. V. M. Bishop, et al., “A manifesto for reproducible science,” *Nature Human Behaviour*, vol. 1, no. 1, 0021, Jan. 2017.
- [6] “USENIX Security ’20 Call for Artifacts.” (2020), [Online]. Available: <https://www.usenix.org/conference/usenixsecurity20/call-for-artifacts> (visited on 02/07/2024).



- [7] "ACM CCS 2023 Call For Artifacts." (2023), [Online]. Available: <https://www.sigmac.org/ccs/CCS2023/call-for-artifacts.html> (visited on 02/07/2024).
- [8] D. Olszewski, A. Lu, C. Stillman, K. Warren, C. Kitroser, A. Pascual, D. Ukirde, K. Butler, and P. Traynor, "Get in Researchers; We're Measuring Reproducibility": A Reproducibility Study of Machine Learning Papers in Tier 1 Security Conferences," in *Proc. 2023 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '23, Copenhagen, Denmark: ACM, 2023, pp. 3433–3459.
- [9] S. Schechter, "Common Pitfalls in Writing about Security and Privacy Human Subjects Experiments, and How to Avoid Them." (2010), [Online]. Available: <https://cups.cs.cmu.edu/soups/2010/howtosoups.pdf>.
- [10] P. B. Stark, "Before reproducibility must come preproducibility," *Nature*, vol. 557, no. 7707, pp. 613–613, May 2018.
- [11] H. E. Plesser, "Reproducibility vs. Replicability: A Brief History of a Confused Terminology," *Frontiers in Neuroinformatics*, vol. 11, 2018.
- [12] L. A. Barba, "Terminologies for Reproducible Research," 2018. arXiv: 1802.03311 [cs.DL].
- [13] ACM. "New Changes to Badging Terminology." (Aug. 24, 2020), [Online]. Available: <https://www.acm.org/publications/badging-terms> (visited on 09/20/2024).
- [14] ACM. "Artifact Review and Badging." version 1.0. (Aug. 24, 2020), [Online]. Available: <https://www.acm.org/publications/policies/artifact-review-badging> (visited on 09/20/2024).
- [15] ACM. "Artifact Review and Badging." version 1.1. (Aug. 24, 2020), [Online]. Available: <https://www.acm.org/publications/policies/artifact-review-and-badging-current> (visited on 09/20/2024).
- [16] SOUPS 2024 PC Chairs. "SOUPS 2024 Call for Papers." (2024), [Online]. Available: <https://www.usenix.org/conference/soups2024/call-for-papers> (visited on 09/20/2024).
- [17] 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 *15th Symposium on Usable Privacy and Security (SOUPS 2019)*, USENIX, Aug. 2019, pp. 117–136.
- [18] K. Pfeiffer, A. Mai, E. Weippl, E. Rader, and K. Krombholz, "Replication: Stories as Informal Lessons about Security," in *18th Symposium on Usable Privacy and Security (SOUPS 2022)*, USENIX, Aug. 2022, pp. 1–18.
- [19] P. Kührtreiber, V. Pak, and D. Reinhardt, "Replication: The Effect of Differential Privacy Communication on German Users' Comprehension and Data Sharing Attitudes," in *18th Symposium on Usable Privacy and Security (SOUPS 2022)*, USENIX, Aug. 2022, pp. 117–134.
- [20] 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 *18th Symposium on Usable Privacy and Security (SOUPS 2022)*, USENIX, Aug. 2022, pp. 367–385.
- [21] H. Aguinis and A. M. Solarino, "Transparency and replicability in qualitative research: The case of interviews with elite informants," *Strategic Management Journal*, vol. 40, no. 8, pp. 1291–1315, Apr. 2019.
- [22] P. T. Sukumar and R. Metoyer, *Replication and Transparency of Qualitative Research from a Constructivist Perspective*, 2019.
- [23] C. G. Begley and J. P. A. Ioannidis, "Reproducibility in science: improving the standard for basic and preclinical research," *Circulation Research*, vol. 116, no. 1, pp. 116–126, Jan. 2015.
- [24] J. R. Stevens, "Replicability and Reproducibility in Comparative Psychology," *Frontiers in Psychology*, vol. 8, 2017.
- [25] J. P. A. Ioannidis, "Why Most Published Research Findings Are False," *PLoS Medicine*, vol. 2, no. 8, pp. 696–701, Aug. 2005.
- [26] M. D. Halbert, "Advancing Reproducibility at the NSF," *Computer*, vol. 55, no. 8, pp. 31–39, 2022.
- [27] S. S. Feger, S. Dallmeier-Tiessen, P. W. Woźniak, and A. Schmidt, "Gamification in Science: A Study of Requirements in the Context of Reproducible Research," in *Proc. 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19, Glasgow, Scotland, UK: ACM, 2019, pp. 1–14.
- [28] P. Dragicevic, Y. Jansen, A. Sarma, M. Kay, and F. Chevalier, "Increasing the Transparency of Research Papers with Explorable Multiverse Analyses," in *Proc. 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19, Glasgow, Scotland, UK: ACM, 2019, pp. 1–15.
- [29] B. Erb, C. Bösch, C. Herbert, F. Kargl, and C. Montag, "Emerging Privacy Issues in Times of Open Science," Jun. 2021.
- [30] A. Chauvette, K. Schick-Makaroff, and A. E. Molzahn, "Open Data in Qualitative Research," *International Journal of Qualitative Methods*, vol. 18, pp. 1–6, 2019.
- [31] B. K. Olorisade, P. Brereton, and P. Andras, "Reproducibility of studies on text mining for citation screening in systematic reviews: Evaluation and checklist," *Journal of Biomedical Informatics*, vol. 73, pp. 1–13, 2017.
- [32] E. Raff, "A Step toward Quantifying Independently Reproducible Machine Learning Research," in *Proc. 33rd International Conference on Neural Information Processing Systems*, Red Hook, NY, USA: Curran Associates Inc., 2019.
- [33] O. E. Gundersen, Y. Gil, and D. W. Aha, "On Reproducible AI: Towards Reproducible Research, Open Science, and Digital Scholarship in AI Publications," *AI Magazine*, vol. 39, no. 3, pp. 56–68, Sep. 2018.
- [34] O. E. Gundersen and S. Kjensmo, "State of the Art: Reproducibility in Artificial Intelligence," 1, vol. 32, Apr. 2018.
- [35] W. Raghupathi, V. Raghupathi, and J. Ren, "Reproducibility in Computing Research: An Empirical Study," *IEEE Access*, vol. 10, pp. 29 207–29 223, 2022.
- [36] J. M. González-Barahona and G. Robles, "On the reproducibility of empirical software engineering studies based on data retrieved from development repositories," *Empirical Software Engineering*, vol. 17, 2012.
- [37] R. P. Reyes, O. Dieste, E. R. Fonseca, and N. Juristo, "Statistical Errors in Software Engineering Experiments: A Preliminary Literature Review," in *Proc. 40th International Conference on Software Engineering*, ser. ICSE '18, Gothenburg, Sweden: ACM, 2018, pp. 1195–1206.
- [38] V. B. Kampenes, T. Dybå, J. E. Hannay, and D. I. K. Sjøberg, "A systematic review of quasi-experiments in software engineering," *Information and Software Technology*, vol. 51, no. 1, pp. 71–82, 2009, Special Section – Most Cited Articles in 2002 and Regular Research Papers.
- [39] J. Miller, "Replicating software engineering experiments: a poisoned chalice or the Holy Grail," *Information and Software Technology*, vol. 47, no. 4, pp. 233–244, 2005.
- [40] J. C. Carver, N. Juristo, M. T. Baldassarre, and S. Vegas, "Replications of software engineering experiments," *Empirical Software Engineering*, vol. 19, 2014.
- [41] N. Ballou, V. R. Warriar, and S. Deterding, "Are You Open? A Content Analysis of Transparency and Openness Guidelines in HCI Journals," in *Proc. 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI '21, Yokohama, Japan: ACM, 2021.
- [42] F. Echterler and M. Häußler, "Open Source, Open Science, and the Replication Crisis in HCI," in *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA '18, Montreal QC, Canada, ACM, 2018, pp. 1–8.
- [43] K. Salehzadeh Niksirat, L. Goswami, P. S. B. Rao, J. Tyler, A. Silacci, S. Aliyu, A. Aebli, C. Wacharamanatham, and M. Cherubini, "Changes in Research Ethics, Openness, and Transparency in Empirical Studies between CHI 2017 and CHI 2022," in *Proc. 2023 CHI Conference on Human Factors in Computing Systems*, ser. CHI '23, ACM, 2023.
- [44] V. Distler, M. Fassel, H. Habib, K. Krombholz, G. Lenzini, C. Lallemand, L. F. Cranor, and V. Koenig, "A Systematic Literature Review of Empirical Methods and Risk Representation in Usable Privacy and Security Research," *ACM Trans. Comput.-Hum. Interact.*, vol. 28, no. 6, Dec. 2021.
- [45] A. A. Hasegawa, D. Inoue, and M. Akiyama, "How WEIRD is Usable Privacy and Security Research?" In *Proc. 33rd USENIX Security Symposium, August 14–16, 2024, Philadelphia, PA, USA*, USENIX, 2024.

- [46] S. S. Feger, P. W. Wozniak, L. Lischke, and A. Schmidt, “‘Yes, I Comply!’: Motivations and Practices around Research Data Management and Reuse across Scientific Fields,” *Proc. ACM Hum.-Comput. Interact.*, vol. 4, no. CSCW2, Oct. 2020.
- [47] B. G. Fitzpatrick, E. Koustova, and Y. Wang, “Getting personal with the ‘reproducibility crisis’: interviews in the animal research community,” *Lab Animal*, vol. 47, 2018.
- [48] Y. Liu, T. Althoff, and J. Heer, “Paths Explored, Paths Omitted, Paths Obscured: Decision Points & Selective Reporting in End-to-End Data Analysis,” in *Proc. 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’20, Honolulu, HI, USA: ACM, 2020, pp. 1–14.
- [49] A. Mobley, S. K. Linder, R. Braeuer, L. M. Ellis, and L. Zwelling, “A survey on data reproducibility in cancer research provides insights into our limited ability to translate findings from the laboratory to the clinic,” *PLoS One*, vol. 8, 2013.
- [50] S. S. Feger, S. Dallmeier-Tiessen, A. Schmidt, and P. W. Woźniak, “Designing for Reproducibility: A Qualitative Study of Challenges and Opportunities in High Energy Physics,” in *Proc. 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’19, Glasgow, Scotland, UK: ACM, 2019, pp. 1–14.
- [51] M. Konkol, C. Kray, and M. Pfeiffer, “Computational reproducibility in geoscientific papers: Insights from a series of studies with geoscientists and a reproduction study,” *International Journal of Geographical Information Science*, vol. 33, no. 2, pp. 408–429, 2019.
- [52] A. Soneji, F. B. Kokulu, C. Rubio-Medrano, T. Bao, R. Wang, Y. Shoshitaishvili, and 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)*, 2022, pp. 1845–1862.
- [53] C. Wacharamanotham, L. Eisenring, S. Haroz, and F. Echter, “Transparency of CHI Research Artifacts: Results of a Self-Reported Survey,” in *Proc. 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’20, , Honolulu, HI, USA, ACM, 2020, pp. 1–14.
- [54] B. Aczel, B. Szaszi, A. Sarafoglou, et al., “A consensus-based transparency checklist,” *Nature Human Behaviour*, vol. 4, no. 1, pp. 4–6, Dec. 2019.
- [55] I. Simerá, D. Moher, J. Hoey, K. F. Schulz, and D. G. Altman, “A catalogue of reporting guidelines for health research,” *European Journal of Clinical Investigation*, vol. 40, no. 1, pp. 35–53, 2010.
- [56] I. Simerá, D. Moher, A. Hirst, J. Hoey, K. F. Schulz, and D. G. Altman, “Transparent and accurate reporting increases reliability, utility, and impact of your research: reporting guidelines and the EQUATOR Network,” *BMC Medicine*, vol. 8, no. 1, Apr. 2010.
- [57] EQUATOR Network. “Enhancing the QUALity and Transparency Of health Research.” (2024), [Online]. Available: <https://www.equator-network.org/> (visited on 01/30/2024).
- [58] K. F. Schulz, D. G. Altman, and D. Moher, “CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials,” *Journal of Clinical Epidemiology*, vol. 63, no. 8, pp. 834–840, 2010.
- [59] B. C. O’Brien, I. B. Harris, T. J. Beckman, D. A. Reed, and D. A. Cook, “Standards for Reporting Qualitative Research,” *Academic Medicine*, vol. 89, no. 9, pp. 1245–1251, Sep. 2014.
- [60] US National Institutes of Health (NIH). “Enhancing Reproducibility through Rigor and Transparency.” (2023), [Online]. Available: <https://grants.nih.gov/policy/reproducibility/index.htm> (visited on 01/22/2024).
- [61] B. A. Nosek, G. Alter, G. C. Banks, et al., “Promoting an open research culture,” *Science*, vol. 348, no. 6242, pp. 1422–1425, 2015.
- [62] Center for Open Science (COS). “TOP Guidelines.” (2015), [Online]. Available: <https://www.cos.io/initiatives/top-guidelines> (visited on 01/30/2024).
- [63] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, et al., “The FAIR Guiding Principles for scientific data management and stewardship,” *Scientific Data*, vol. 3, no. 1, Mar. 2016.
- [64] “FAIR Principles.” (2016), [Online]. Available: <https://www.go-fair.org/fair-principles/> (visited on 01/30/2024).
- [65] A. J. Ko, T. D. LaToza, and M. M. Burnett, “A practical guide to controlled experiments of software engineering tools with human participants,” *Empirical Software Engineering*, vol. 20, no. 1, pp. 110–141, Sep. 2013.
- [66] A. Jedlitschka and D. Pfahl, “Reporting guidelines for controlled experiments in software engineering,” in *2005 International Symposium on Empirical Software Engineering*, 2005.
- [67] A. Jedlitschka, M. Ciolkowski, and D. Pfahl, “Reporting Experiments in Software Engineering,” in *Guide to Advanced Empirical Software Engineering*, F. Shull, J. Singer, and D. I. K. Sjöberg, Eds. London: Springer, 2008, pp. 201–228.
- [68] F. J. Shull, J. C. Carver, S. Vegas, and N. Juristo, “The role of replications in Empirical Software Engineering,” *Empirical Software Engineering*, vol. 13, no. 2, pp. 211–218, Jan. 2008.
- [69] P. Ralph, S. Baites, D. Bianculli, et al., *Empirical Standards for Software Engineering Research*, 2020. arXiv: 2010.03525 [cs.SE].
- [70] “Empirical Standards.” (2020), [Online]. Available: <https://www2.sigsoft.org/EmpiricalStandards/about/> (visited on 01/30/2024).
- [71] L. L. Chuang and U. Pfeil, “Transparency and Openness Promotion Guidelines for HCI,” in *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA ’18, ACM, 2018, pp. 1–4.
- [72] C. Wacharamanotham, M. Kay, S. Haroz, S. Guha, and P. Dragicevic, “Special Interest Group on Transparent Statistics Guidelines,” in *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA ’18, ACM, 2018, pp. 1–4.
- [73] “Security Research Artifacts.” (2024), [Online]. Available: <https://secartifacts.github.io/> (visited on 01/30/2024).
- [74] J. Lazar, J. H. Feng, and H. Hochheiser, *Research Methods in Human-Computer Interaction*. Wiley, 2010.
- [75] P. J. P. Cormac McGrath and M. Liljedahl, “Twelve tips for conducting qualitative research interviews,” *Medical Teacher*, vol. 41, no. 9, pp. 1002–1006, 2019.
- [76] B. Saunders, J. Sim, T. Kingstone, S. Baker, J. Waterfield, B. Bartlam, H. Burroughs, and C. Jinks, “Saturation in qualitative research: exploring its conceptualization and operationalization,” *Quality & Quantity*, vol. 52, pp. 1893–1907, 2017.
- [77] A.-M. Ortloff, M. Fassl, A. Ponticello, F. Martius, A. Mertens, K. Kromholz, and 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 *Proc. CHI Conference on Human Factors in Computing Systems*, ser. CHI ’23, New York, NY, USA: ACM, 2023.
- [78] “Human-Made Transcription.” (2024), [Online]. Available: <https://www.amberscript.com/en/products/manual-transcription/> (visited on 02/08/2024).
- [79] V. Braun and V. Clarke, “Using thematic analysis in psychology,” *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, 2006.
- [80] N. McDonald, S. Schoenebeck, and A. Forte, “Reliability and Inter-Rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice,” *ACM on Human-Computer Interaction*, vol. 3, no. CSCW, 72, pp. 1–23, 2019.
- [81] E. Kenneally and D. Dittrich, “The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research,” U.S. Department of Homeland Security, Tech. Rep., Aug. 2012.
- [82] E. Kenneally and D. Dittrich, “Applying Ethical Principles to Information and Communication Technology Research: A Companion to the Menlo Report,” U.S. Department of Homeland Security, Tech. Rep., Oct. 2013.
- [83] M. Bailey, D. Dittrich, E. Kenneally, and D. Maughan, “The Menlo Report,” *IEEE Security & Privacy*, vol. 10, no. 2, pp. 71–75, 2012.
- [84] ACM Publications Board. “ACM Publications Policy on Research Involving Human Participants and Subjects.” (2021), [Online]. Available: <https://www.acm.org/publications/policies/research-involving-human-participants-and-subjects> (visited on 01/30/2024).

- [85] APA. “Preregistration.” (Jan. 2021), [Online]. Available: <https://www.apa.org/pubs/journals/resources/preregistration> (visited on 02/02/2024).
- [86] V. Huser and J. J. Cimino, “Evaluating adherence to the International Committee of Medical Journal Editors’ policy of mandatory, timely clinical trial registration,” *Journal of the American Medical Informatics Association*, vol. 20, no. e1, e169–e174, Feb. 2013.
- [87] “NeurIPS Paper Checklist Guidelines.” (2023), [Online]. Available: <https://neurips.cc/public/guides/PaperChecklist> (visited on 02/05/2024).
- [88] A. Tong, P. Sainsbury, and J. Craig, “Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups,” *International Journal for Quality in Health Care*, vol. 19, no. 6, pp. 349–357, Sep. 2007.
- [89] C. Anderson, “Presenting and Evaluating Qualitative Research,” *American Journal of Pharmaceutical Education*, vol. 74, no. 8, p. 141, Oct. 2010.
- [90] L. Busetto, W. Wick, and C. Gumbinger, “How to use and assess qualitative research methods,” *Neurological Research and Practice*, vol. 2, no. 1, May 2020.
- [91] I. Hrynaskiewicz, M. L. Norton, A. J. Vickers, and D. G. Altman, “Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers,” *Trials*, vol. 11, no. 1, Jan. 2010.
- [92] T. Holz and A. Oprea. “IEEE S&P’21 Program Committee Statement Regarding The “Hypocrite Commits” Paper.” (May 6, 2021), [Online]. Available: [https://www.ieee-security.org/TC/SP2021/downloads/2021\\_PC\\_Statement.pdf](https://www.ieee-security.org/TC/SP2021/downloads/2021_PC_Statement.pdf) (visited on 02/02/2024).
- [93] Y. Zhang, M. Liu, M. Zhang, C. Lu, and H. Duan, “Ethics in Security Research: Visions, Reality, and Paths Forward,” in *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 2022, pp. 538–545.
- [94] “Best of CHI 2013.” (2013), [Online]. Available: <https://chi2013.acm.org/program/best-of-chi/> (visited on 02/02/2024).
- [95] ACM. “Ceasing Print Publication of ACM Journals and Transactions.” (2023), [Online]. Available: <https://www.acm.org/publications/ceasing-print> (visited on 02/02/2024).
- [96] “Proceeding on Privacy Enhancing Technologies Symposium - Information for Authors.” (2024), [Online]. Available: <https://petsymposium.org/authors.php#submission-guidelines> (visited on 02/02/2024).
- [97] ICPSR. “Guide to Social Science Data Preparation and Archiving, Best Practice Throughout the Data Life Cycle: 6th Edition.” (2020), [Online]. Available: <https://www.icpsr.umich.edu/files/deposit/dataprep.pdf> (visited on 02/08/2024).
- [98] “dblp.” (2024), [Online]. Available: <https://dblp.org/> (visited on 05/14/2024).
- [99] U.S. Department of Health & Human Services. “2018 Common Rule.” 45 CFR 46. (2018), [Online]. Available: <https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/revised-common-rule-regulatory-text/index.html>.
- [100] J. R. Landis and G. G. Koch, “The Measurement of Observer Agreement for Categorical Data,” *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977.

## Appendix A. UPS Paper Dataset

We utilized bibliographic data from DBLP [98] to compile a dataset of UPS conference papers between 2018 and 2022. We considered the proceedings of *IEEE S&P*, *ACM CCS*, *USENIX Security*, *NDSS*, *IEEE EuroS&P*, *PETS/PoPETs*, *ACM CHI*, *ACM CSCW*, *SOUPS*, *EuroUSEC<sup>2</sup>*, *IEEE/ACM ICSE*, and *ACM WWW*, with a total of 10,827 published full papers. We identified UPS papers in a manual coding process based on the following definition:

**Definition:** We consider a publication a usable security and privacy paper, if it (1) covers the topics security and/or privacy, and (2) is human subjects research. We follow the human subjects research definition of the *2018 Common Rule* that can be summarized as: “Research involving a living individual about whom data or biospecimens are obtained/used/studied/analyzed through interaction/intervention, or identifiable, private information is used/studied/analyzed/generated” [99].

Three authors first coded a random sample of 100 papers independently, achieving an *almost perfect* inter-rater-reliability (Fleiss’ kappa = 0.89) [100]. Given the high agreement and the decision requiring only little interpretation [80], the three authors split coding of the remaining dataset. We identified 727 UPS papers, which we used to sample authors for our recruitment. We provide the UPS paper dataset with our other artifacts as described in the Availability section.

2. For EuroUSEC we consider only the 2021 and 2022 proceedings, as EuroUSEC turned from workshop to symposium in 2021.

## Appendix B. Meta-Review

The following meta-review was prepared by the program committee for the 2025 IEEE Symposium on Security and Privacy (S&P) as part of the review process as detailed in the call for papers.

### B.1. Summary

This paper qualitatively studies the perspectives of the Usable Privacy and Security (UPS) researchers about practices and challenges related to transparency in the context of the replicability/reproducibility crisis.

### B.2. Scientific Contributions

- Addresses a Long-Known Issue
- Provides a Valuable Step Forward in an Established Field

### B.3. Reasons for Acceptance

As the long-known reproducibility crisis may also affect the UPS research, it is essential to understand how transparency, replicability, and reproducibility are understood and practiced within the community and the challenges experienced/perceived by practitioners. Through a qualitative study, the paper identifies that while the UPS community desires and values transparency and replicability, the understanding and practices vary within the community. Thus, this paper highlights the need to clarify these concepts in this established field so that UPS research can be further improved while ensuring transparency.

### B.4. Noteworthy Concerns

One of the reviewers had three noteworthy concerns with the paper:

- 1) **Defining and then exploring transparency:** The study should have provided the participants with a definition of transparency at the start of the interviews, and then asked participants whether they agreed with the definition. If not, the study should have explored how the definition should be extended/modified and refined in several dimensions to contextualize it for UPS research. As the study did not do so, it failed to obtain in-depth insights about transparency.
- 2) **Unsurprising results:** One of the results, emphasized in the paper rebuttal, is that participants largely shared the same understanding of the meaning of transparency. This is not surprising, i.e., the reviewer believes that one would not need a study to conclude this. It would have been surprising if they had a different understanding. If the participants had been provided a clear definition of transparency beforehand (as stated in #1), differences would have emerged, as they do when one looks

deeply into the dimensions of transparency. Therefore, the reviewer feels that the results of the study are quite obvious concerning the notion of transparency for UPS.

- 3) **Replication:** From the paper and subsequent discussion, the reviewer did not see a clear notion or articulation of replication in the paper. For instance, in other disciplines (for example psychology), papers are withdrawn because of data/other errors in studies. In this context, the paper does not define a clear goal for replication of UPS. Moreover, it does not address a key question: If a replication study gets different results from the original work, can one determine which study is correct and which is not? Given that it is not possible to exactly replicate a study when human subjects are involved (especially if participants are anonymized due to privacy concerns), it is unclear how one be able to reach actionable conclusions. To summarize, the reviewer observes that the paper does not address key concerns regarding replication, and believes that its findings are obvious in the other aspect, i.e., transparency.

However, the other reviewers concluded that the study offered valuable insights into the implicit understanding of transparency and replication within the UPS community. Therefore, this study will be helpful for future efforts toward contextualizing transparency, replication, reproducibility, and establishing relevant guidelines.

## Appendix C. Response to the Meta-Review

We provide the following response to the one reviewer's noteworthy concerns:

- 1) **Defining and then exploring transparency:** Providing participants with a definition of transparency would have been another study approach. We deliberately decided not to provide such a definition for several reasons. We sought to explore participants' perceptions, experiences, and opinions, and did not want to induce bias or prime participants with a formal transparency definition or asking whether they agree with it (cf. Section 4.2). This is common in qualitative interviews to elicit and understand participants' interpretations and definitions of a given concept, and thus be able to analyze the extent to which interviewees' understanding of a concept (here transparency) is consistent or substantially differs among participants. Moreover, presenting a definition initially would limit the study to be more confirmatory, while our study goal was exploratory. Measuring definition agreement might also be done in a large-scale survey, and could be a potential future work following our insights.
- 2) **Unsurprising results:** First, we note that the meaning/definition of transparency is not our only result. Second, while not all results may be "surprising," e.g., because some results reflect anecdotal knowledge of individual researchers, we disagree that our results do not provide novel insights. Generally, we argue that a

scientific contribution does not need to be “surprising” to be valuable. Considering that this is meta-research within the UPS community, it is highly unlikely to find only “surprising” results, but to also confirm and surface implicit norms and practices as our study did (which this reviewer considered “unsurprising” knowledge). Our study is the first systematic analysis of transparency within the (U)PS community, and therefore a novel contribution. One of the other reviewers fittingly summarized this: “*Although some of the observations are pretty well known, it is still nice to see them being validated and generating new discussion.*”

From a meta perspective with our interview insights, we think that the reviewer’s concern is a good example of favoring positive/surprising results which could lead to a publication bias—which was a concern in our interviews. Many participants perceived a bias against “unsurprising” but validating results, i.e., null results (cf. Section 5.5).

3) **Replication:** After discussions in the rebuttal phase, we added Section 2 to explain the terms transparency, replicability, and reproducibility as well as their relationship and differences. We note that this paper is not primarily about replication (in the sense of actually obtaining the same results), but transparency (in the sense of reporting research, e.g., so that others can *attempt* a replication). We report on the former, as participants brought replication up and related both concepts.

We consider the meta-review’s question “*If a replication study gets different results from the original work, can one determine which study is correct [...]?*” out of our study’s scope, which could be explored for (U)PS in future work. Apart from that, we note that other disciplines approach this question with methods like *meta-analysis*, e.g., in medicine, and that revealing such inconsistencies is one of the purposes of replication.