Understanding Risks of Privacy Theater with Differential Privacy

MARY ANNE SMART, University of California San Diego, USA
DHRUV SOOD, University of California San Diego, USA
KRISTEN VACCARO, University of California San Diego, USA

Differential privacy is one of the most popular technologies in the growing area of privacy-conscious data analytics. But differential privacy, along with other privacy-enhancing technologies, may enable privacy theater. In implementations of differential privacy, certain algorithm parameters control the tradeoff between privacy protection for individuals and utility for the data collector; thus, data collectors who do not provide transparency into these parameters may obscure the limited protection offered by their implementation. Through large-scale online surveys, we investigate whether explanations of differential privacy that hide important information about algorithm parameters persuade users to share more browser history data. Surprisingly, we find that the explanations have little effect on individuals' willingness to share data. In fact, most people make up their minds about whether to share before they even learn about the privacy protection.

CCS Concepts: • **Human-centered computing** → **Human computer interaction (HCI)**; User studies; • **Security and privacy** → **Human and societal aspects of security and privacy**; **Usability in security and privacy**.

Additional Key Words and Phrases: privacy theater, differential privacy, human-centered privacy

ACM Reference Format:
Mary Anne Smart, Dhruv Sood, and Kristen Vaccaro. 2022. Understanding Risks of Privacy Theater with Differential Privacy. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 342 (November 2022), 24 pages. https://doi.org/10.1145/3555762

1 INTRODUCTION

In recent years, the HCI community has devoted increasing attention toward the issues of mass data collection and surveillance capitalism [29, 83, 88]. The role of privacy-enhancing technologies (PETs) within this landscape is complicated. On the one hand, PETs can offer protection from and resistance against harmful data collection practices [10]. On the other hand, some PETs may actually normalize surveillance, increase the power of data collectors, de-politicize surveillance issues by reframing them as technological puzzles, or otherwise distort political discourse around surveillance issues [33, 68, 77, 88]. In addition, these PETs may encourage *privacy theater*, where they provide the "*feeling of improved privacy while doing little or nothing to actually improve privacy*" [46]. We investigate this possibility using one popular PET: differential privacy.

Differential privacy has become one of the most widely used tools for privacy-conscious data analytics, with deployments across industry and government agencies (e.g., Google [22], Apple [6], US Census Bureau [1]). Differential privacy allows these organizations to collect data while protecting privacy by adding small amounts of statistical noise to the data that people share [19].

Authors' addresses: Mary Anne Smart, msmart@ucsd.edu, University of California San Diego, San Diego, California, USA; Dhruv Sood, dhsood@ucsd.edu, University of California San Diego, San Diego, California, USA; Kristen Vaccaro, kv@ucsd.edu, University of California San Diego, San Diego, California, USA.

2573-0142/2022/11-ART342
https://doi.org/10.1145/3555762

Importantly, the privacy protection provided by differential privacy is highly dependent on the setting of certain algorithm parameters, which control how much noise is added. However, there is an inherent tradeoff between the noise added (privacy protection provided) and the utility of the data for the organization. As a result, there is a risk that companies may misrepresent the actual privacy benefits afforded by differential privacy to persuade users to share more information. By understanding how users respond to different kinds of explanations of differential privacy, we can better understand whether users are likely to be harmed by explanations that obscure these tradeoffs.

Recent work has begun to probe users' understanding of differential privacy. For example, recent work studied how different explanations of differential privacy influenced users' (planned) willingness to share their data [87]. These explanations did not describe the role of algorithm parameters to users, because the authors argued that users would struggle to understand how parameter settings would affect their privacy. A different study of differential privacy, however, provided transparency into the amount of noise added to participants' data. This transparency increased participants' comfort, understanding, and trust in the security of the differentially private mechanism [11]. In this work, we expand on these efforts, developing explanations of differential privacy that convey information about algorithm parameters and investigating whether incomplete explanations of differential privacy may mislead users.

Through multiple large-scale, online surveys, we study how different explanations of differential privacy influence participants' understanding and behavior when asked to share their browser history data. Drawing and expanding on prior work, the first experiment aims to generate a "best possible" explanation for differential privacy, varying 1) whether the explanation is text-only or includes a visual component, and 2) whether the explanation focuses on the process or outcome of the differentially private mechanism. The second experiment tests for potential *privacy theater*, pitting the best-performing explanation against standard industry explanations, and observing whether industry explanations obscure risks of poor privacy settings. Unlike prior work [87], we do not assume that more willingness-to-share is always preferable. Instead, we hypothesize that good explanations will decrease willingness-to-share when privacy protection is poor.

Our results reveal a number of surprising findings. First, we find that participants are overconfident in their understanding; although they feel that they have understood the explanations well, they perform poorly on comprehension questions. In particular, a surprisingly large number of participants do not understand that differential privacy offers protection from a wide range of adversaries—not just from hackers. This suggests that participants find the nature of the protection offered by differential privacy to be counterintuitive. The second surprising finding is that the explanations of differential privacy have little effect on individuals' willingness to share browser history data. Only a small fraction of participants—less than 25%—actually change their minds after reading about the privacy protection; most participants have already decided whether or not to share their data. Our qualitative results help explain the reasons for this behavior. Many participants list other reasons that drive their decision making around disclosing browser data, such as trust (or distrust) in the research team. Many participants simply feel that since they have "nothing to hide," they may as well share their information. For these participants, promises of privacy protection are largely irrelevant.

2 RELATED WORK

Attempts to develop good explanations of differential privacy face a key challenge: it is difficult to explain how specific algorithm parameters affect privacy risk. In fact, understanding how to best set the privacy parameter remains a challenge even for experts [1, 41]. This section begins

with some necessary background about differential privacy. It continues with an overview of the relevant prior work on privacy-related communication.

2.1 Differential Privacy

Differential privacy has received a great deal of attention in academia, industry, and government, for its potential to empower data analytics with provable privacy guarantees. The goal of differential privacy is to allow data analysts to learn from *aggregate* statistics while limiting the leakage of information that is specific to any *individual*. Differentially private mechanisms add noise to data to accomplish this. However, as more noise is added, the data becomes less useful for analysis.

2.1.1 Formal Definition. Two common variants exist: central differential privacy and local differential privacy (LDP). In the central model, a central agent is responsible for storing the raw data and adding noise. In the local model, individuals add noise to their own data before sending it to the data collector. Thus, the local model offers stronger privacy protection, since it does not require trust in a central agent. Local differential privacy will be our focus in this paper.

The formal definition of local differential privacy reveals the importance of a parameter for how much privacy protection is provided. It states [19] that a randomized algorithm A is $\boldsymbol{\varepsilon}$**-LDP** if and only if for any inputs u, v and any outputs $y \in Range(A)$:

$$\ln\left(\frac{P(A(u) = y)}{P(A(v) = y)}\right) < \varepsilon$$

Since this definition requires the inequality to hold for all values of u, v, and y, it offers a kind of worst-case guarantee; replacing input u with any other input v will not change the probability distribution over algorithm outputs much, provided that the privacy parameter ε is small. In effect, when ε is small, anyone looking at an individual datapoint will be uncertain about its *true* value. Note that this privacy parameter[1], which we discuss in more detail below, is an essential factor in determining just how much privacy is offered.

2.1.2 Tradeoffs. To illustrate the tradeoff the privacy parameter makes between privacy and utility, we describe a popular survey technique that satisfies LDP[2]: the randomized response technique. Suppose a researcher wants to estimate the proportion of students at a particular school who use illegal drugs [32]. Since "*Have you ever used illegal drugs?*" is a sensitive question that many students may be reluctant to answer, the researcher gives students the following instructions. Privately flip a coin. If the coin toss is heads, write down your true answer. Otherwise, flip the coin again. If the second coin toss is heads, answer *yes*; if tails, answer *no*. This procedure offers a form of privacy to students; someone who saw a student's response would not know whether it was the true response or not. In fact, this mechanism is ln(3)-LDP [22]. Nevertheless, these answers allow an estimate of the proportion of students that use illegal drugs. If out of n students surveyed, y respond *yes*, then we can estimate that the true proportion of students who use illegal drugs is about $2(\frac{y}{n} - \frac{1}{4})$.

The privacy parameter ε controls the tradeoff between privacy and utility. Suppose that instead of flipping a coin, students roll a six-sided die. The student is instructed to report their true answer unless they roll a 1, in which case they flip a coin and report *yes* for heads and *no* for tails. This provides a more accurate estimate of how many students use illegal drugs, however, it reduces the privacy offered to students. Someone who sees a student's response to the question has more reason to believe that the response is true. This mechanism no longer satisfies ln(3)-LDP, but it

[1]The privacy parameter ε is often referred to as the privacy *budget*. In other variants of differential privacy, there may be multiple privacy parameters, but in our setting, we can unambiguously refer to the privacy budget as *the* privacy parameter.
[2]Though RRT was invented before the formalization of differential privacy [85]

does satisfy ln(11)-LDP. This example illustrates how increasing the privacy parameter ϵ increases the usefulness of the data for analysis, but weakens the privacy guarantee.

2.1.3 Industry Deployments. This privacy parameter setting has become an area of contention for industry deployments. One early use of LDP was Google's deployment of RAPPOR (which builds upon randomized response [22]), to collect information about certain Chrome users' browser settings. Apple followed suit, deploying its own versions of LDP to collect data for a variety of applications including discovery of new words, discovery of popular emojis, and analysis of memory usage in the Safari browser [6]. But Apple faced criticism for its initial decision not to publish the privacy parameter ε. When researchers reverse-engineered Apple's implementation of LDP on MacOS 10.2, they found the privacy loss to be "*significantly higher than what is commonly considered reasonable in academic literature*" [76]. In fact, this lack of transparency has been recognized as such a significant issue that one of the inventors of differential privacy called for the creation of an "Epsilon Registry," for firms to report implementation details such as the choice of privacy parameter ε. The underlying concern is that *"when ε is large it can [...] allow for a form of privacy theatre"* while offering little privacy protection [20].

2.2 Communicating Privacy

Researchers have extensively studied the questions of how privacy-related communication affects user behavior [2, 75] and how to communicate privacy risks effectively [8, 21, 25, 30, 31, 44, 45]. Previous studies have pointed out the shortcomings of traditional privacy policies [59, 80]. In theory, privacy policies could help people make informed decisions about their privacy. In practice, however, this is rarely the case [55]. In fact, privacy policies can actually be misleading. For example, people sometimes incorrectly interpret the mere presence of a privacy policy as an assurance that the website does not share users' data with third-parties [78]. Similar to this prior work, we are concerned that ineffective communication could give users a false sense of security.

Recently, these efforts have extended to how to communicate with users about differential privacy [11, 43, 87]. The study most closely related to ours is that of Xiong et al., which examines how different explanations of differential privacy shape users' willingness to share their data [87]. However, this work evaluates the quality of explanations in terms of users' resulting willingness to share. They state, for example, that "*when definitions of DP and LDP were communicated . . . participants increased their data disclosure for high-sensitive information, suggesting a positive effect of communicating differential privacy to laypeople.*" In this work, we explore the potential risks involved in this framing and investigate whether misleading explanations can lead to increased data sharing when little actual privacy protection is provided.

Several previous studies have investigated factors that can shift users' willingness to share their data in different contexts. For example, iPhone users who are asked to grant certain permissions to an app are more likely to grant the permissions when the app provides an explanation of why the requested permissions are necessary [75]. A more nefarious example can be found in the study of phishing. Phishing emails seek to trick individuals into disclosing valuable information, such as banking credentials. Phishing emails that are information-rich tend to be more successful [27]. In this work, we seek to understand the malleability of participants' existing data sharing preferences and to what extent participants' behavior can be shaped by explanations of differential privacy.

3 EXPERIMENTAL DESIGN

Differential privacy is an approach for providing privacy by adding noise to the data users provide, resulting in a tradeoff between privacy protection (which increases as noise is added) and the usefulness of the data (which decreases as noise is added). This study develops explanations of

Table 1. Example Table

#	Website	Visited?
1	example.com	YES
2	example.org	NO
...	...	...
200	example.net	YES

Table 2. Settings of Privacy Parameter

Flip Probability	ε
1%	920
25%	220
49%	9

differential privacy that make the necessary trade-offs between privacy and utility explicit. The first experiment identifies the explanation that is most effective at communicating this information to users. This experiment tests four explanations' effectiveness at improving users' perceived and actual understanding of differential privacy. The second experiment measures how well explanations can convey the risks of poor settings of algorithm parameters and whether better understanding can induce more appropriate behavior (specifically, lower willingness to share data).

3.1 Internet Browsing Histories

The setting used for these experiments was collecting internet browsing history data. Prior experiments on explaining differential privacy have assessed *hypothetical* willingness to share data [87]. Unlike this prior work, our experiment seeks to have participants *actually decide* whether or not to share their private data. Browsing history data is an appropriate choice since many people consider browsing history data sensitive [65, 70]. Furthermore, many large-scale industry deployments of differential privacy have involved the collection of browser-related data [6, 22]. It is also plausible that the experiment could access participants' actual data since the experiment is conducted through an online survey.

Participants were told that we were interested in collecting internet browsing histories, focused on a list of 200 websites of interest. Using participants' browsers, an automated script would collect all websites visited over the past 12 months. From this list, we could construct a table showing which of the 200 websites had been visited (Table 1). To satisfy differential privacy, the data collection would randomly change some answers. Participants were told that only this modified data would be shared with the researchers; the original data would never leave the participants' browser. The explanation provided to participants is included in Table 3.

The percentage of changed answers is directly related to the privacy parameter ε (Table 2). By varying the percentage of answers that are changed, we vary the privacy parameter. As a result, the privacy parameter can be changed as an experimental condition.

While participants were told that they were deciding whether to share data, no browsing data was actually collected. Participants were debriefed on this deception at the end of the experiment. We did not deceive participants about our identity as researchers. Participants were explicitly informed that any data they shared would be shared with an academic research team. Our Institutional Review Board determined that the protocol was exempt from full IRB review.

3.2 Experiment 1: Developing Good Explanations of Differential Privacy

The first experiment evaluated which explanations of differential privacy would be easy for participants to understand. Prior work found that explanations that covered the implications of local differential privacy—rather than the definition or process of adding noise to data—improved participants' comprehension [87]. And work on descriptions of encryption found that result-oriented descriptions led to greater feelings of security than process-oriented descriptions [16]. Nevertheless,

explanations that fail to explain the underlying processes may confuse users, and other authors have argued that *"visibility of system behavior"* is essential for encouraging good decision-making around privacy and security [17]. To test which is most useful to increase understanding of differential privacy, we developed two explanation texts, one of which focuses on the *outcome* of the privacy-preserving process and the other of which focuses on the *process* itself. This led to our first hypothesis[3]:

H1 *Process vs. Outcome:* Explanations focused on outcomes will increase understanding more than explanations focused on process.

Previous work has also found visual aids to be helpful in aiding understanding of differential privacy [11] and in understanding other privacy-related issues [54, 82, 86]. Visualizations have also played an important role in attempts to increase the explainability and interpretability of machine learning models [50, 71]. However, other work has found that people with limited statistics background often struggle to interpret visualizations related to probability distributions [38]. This presents a challenge, since it is important that our explanations convey information about the randomness inherent to differential privacy. These findings from previous work led to our second hypothesis:

H2 *Visual explanations:* Explanations providing a visual component will improve understanding compared to a text-only version of the explanation.

3.2.1 Experimental Conditions. To test these hypotheses, we use a 2^2 factorial design. The two factors are content and medium. Each has two levels; the content provides process- and outcome-focused content, and the medium provides either a text-only or text+visual explanation.

The four explanations are roughly matched for length (160-163 words) and reading level (6-8 on the Flesch-Kincaid grade level scale [47]). All also convey the same key information: 1) the kind of changes that would be made to protect privacy, i.e., that the record of whether particular websites had been visited would be changed, and that these changes would be made at random, 2) information related to the privacy parameter, i.e., how many changes would be made, and 3) the fact that the researchers would only see the altered data, and that as a result the researchers would be uncertain about which websites were actually visited. Four pilot testers—including one differential privacy expert—were recruited to evaluate the explanations on the basis of correctness, interpretability, and clarity. Discussions with these pilot testers revealed a desire to understand how the collected data could be used after undergoing the randomized privacy-preserving process; a final sentence clarifies this. Table 3 provides the explanations.

One challenge in developing visual explanations of differential privacy is how to convey the randomness inherent to all differentially private mechanisms. Our visualizations belong to the broader class of hypothetical outcome plots, which use animation to simulate the outcomes of multiple random draws [60]. The visual components, which take the form of animated GIFs, can be found in the supplementary materials.

3.3 Experiment 2: Measuring Privacy Theater

The second experiment examines how different explanations influence participant behavior—particularly willingness to share data. Of particular concern is how users decide whether to share data when presented with explanations that provide little to no transparency into the setting of

[3] All four hypotheses were preregistered with the Open Science Framework.

Table 3. Explanations for Experiment 1

Process	Outcome
We will collect information about your internet browsing history using a method designed to protect privacy. The method provides local differential privacy. We have a list of 200 websites that interest us. First, the method will look at your browsing history from the past year. It will ignore any websites that are not on our list of interest. Then it will make a table. The table will show which of the 200 websites you visited. The table would look something like this: (see Table 1).	
The method will randomly select some rows in the table to change. Each row has a 49% chance of being selected to change. For these rows, the method will change the YES answers to NO and the NO answers to YES. We will not know which rows were selected. These changes will happen before you send us your information.	The method will have changed your information before you send it. For most people, about 98 of the 200 YES or NO answers will have changed. The changes will be different for each person. We will never have access to the original table. We also will not know what parts of your browsing history were changed.
Differential privacy changes each individual's information. But since we collect information from many people, we can still see overall patterns that interest us.	

the privacy parameter. We refer to these explanations that omit any discussion of the privacy parameter and its implications as *low-transparency* explanations; similarly, we refer to explanations that *do* discuss the implications of the privacy parameter as *high-transparency* explanations. If users are more willing to share data when given a low-transparency explanation than when given a high-transparency explanation, then data collectors may be incentivized to use low-transparency explanations. Furthermore, data collectors may be tempted to set inappropriately large values for the privacy parameter and thus abuse differential privacy as privacy theater.

Prior work suggests that explaining differential privacy's protections can make people more willing to share sensitive data [28, 87]. Highlighting the particular benefits of *local* differential privacy has also been found to increase willingness to share [87]. So when the privacy protection offered is strong— i.e., the privacy parameter is small—highlighting this fact will likely increase willingness to share. We hypothesize that in the strong privacy setting (i.e. small ε), participants would be more willing to share their data when presented with a more transparent explanation that better highlights the strength of the privacy guarantee.

H3 *Strong Privacy Setting:* When the privacy protection is strong, users provided with a high-transparency explanation will be more willing to share data than those provided with a low-transparency explanation.

Prior work demonstrates that even explanations that omit any discussion of the privacy parameter can increase participants' willingness to share data [87]. In the weak privacy setting (i.e. large ε), an explanation of the privacy parameter might reveal the weakness of the privacy protection being offered, such that the increased willingness to share observed in [87] would disappear. Indeed, previous work on the randomized response technique (RRT)—the simplest variant of differential

privacy—suggests that this is likely to happen, since it has been shown that participants' trust, comfort, and level of perceived protection fall as the privacy parameter increases [11, 72]. We hypothesize that in the weak privacy setting (i.e. large ε), participants would be less likely to share their information when given more transparent explanations.

H4 *Weak Privacy Setting:* When the privacy protection is poor, users provided with a low-transparency explanation will be more willing to share data than those provided with a high-transparency explanation.

3.3.1 Experimental Conditions. To test these hypotheses, we use a 3^2 factorial design. The two factors are transparency and privacy parameter setting. Each has three levels; the privacy parameter setting levels provide high, medium, and low privacy protection, and the explanations provide high, medium, and low transparency into the privacy parameter.

The first factor is the privacy parameter setting. We have a low privacy setting ($\varepsilon = 920$), a medium privacy setting ($\varepsilon = 220$), and a high privacy setting ($\varepsilon = 9$). In general, the low privacy setting changes very few of users' true responses, while the high privacy setting changes many of the true responses. Therefore, as the level of privacy increases, the data becomes less useful for the data collector. Table 2 indicates exactly how these ε values translate into privacy protection.

The value of the privacy parameter in the high privacy setting is quite close to what has been used for certain industry deployments [6]. Nevertheless, future work could consider even smaller values, such as $\varepsilon < 1$, for even stronger privacy guarantees. The privacy offered in the low privacy

Table 4. Low, Medium, and High Transparency Explanations for Experiment 2. All explanations are for the high privacy setting. The low transparency explanation conveys no information about the privacy parameter.

Low *(Uber)*	Differential privacy is a formal definition of privacy and is widely recognized by industry experts as providing strong and robust privacy assurances for individuals. In short, differential privacy allows general statistical analysis without revealing information about a particular individual in the data. Differential privacy provides an extra layer of protection against re-identification attacks as well as attacks using auxiliary data.
Med *(Apple)*	Differential privacy transforms the information shared with us before it ever leaves the user's device such that we can never reproduce the true data. The differential privacy technology used is rooted in the idea that statistical noise that is slightly biased can mask a user's individual data before it is shared with us. If many people are submitting the same data, the noise that has been added can average out over large numbers of data points, and we can see meaningful information emerge. Our differential privacy implementation incorporates the concept of a privacy budget (quantified by the parameter epsilon). We use a privacy budget with epsilon of 9.
High *(Ours)*	The method will have changed your information before you send it. For most people, about 98 of the 200 YES or NO answers will have changed. The changes will be different for each person. We will never have access to the original table. We also will not know what parts of your browsing history were changed. Differential privacy changes each individual's information. But since we collect information from many people, we can still see overall patterns that interest us.

setting, however, is much worse than what would be considered reasonable in practice—in fact, even the parameter value for the medium privacy setting offers relatively weak privacy guarantees [76]. This extreme value for the low privacy setting makes it clear that differential privacy in this case offers only privacy theater rather than any meaningful privacy protection.

The second factor is transparency. The explanations provide high, medium, and low transparency into the privacy parameter (Table 4). The high transparency explanation was the outcome-focused explanation from Experiment 1, as it directly conveyed the implications of the privacy parameter for this data. The medium and low transparency explanations are drawn from real industry explanations of differential privacy. The low-transparency explanation is an excerpt from an Uber blog post[4]. This explanation does not mention the privacy parameter. The medium transparency explanation is one that provides information on the privacy parameter, but in a way that reveals little useful information to users. This explanation is an adapted excerpt from a larger document in which Apple explains its implementation of local differential privacy [6]. Interestingly, this document gives the exact ε values for Apple's implementations of differential privacy; however, the document does not provide much context to help readers interpret the numbers. Both explanations were found to be effective in persuading users to share sensitive information in prior work [87].

The explanations were roughly matched for word count (60-108 words). But rather than controlling for reading level or content, we left the industry language intact. Both use technical language and jargon, which may be an intentional feature. Some scholars have noted that jargon-laden privacy policies actively discourage users from engaging with the information they contain [18, 63]; it is possible that the use of jargon in explanations of differential privacy could play a similar role. A text that leans heavily on technical jargon may give the sense that there exists "*some sort of crypto-magic to protect people from data misuse*" without actually providing an explanation that is accessible for the average user [68]. Similarly, the Uber explanation makes an appeal to authority—referencing "industry experts"—which may be designed to exploit authority bias [42].

3.4 Measures

Both experiments used the same measures for dependent variables: perceived understanding, actual understanding, and willingness to share. Similarly, both included measures for covariates including: demographics, general privacy concerns, sensitivity and internet browser usage. The full survey text is included in the supplementary materials.

3.4.1 Understanding. Drawing on prior work [87], we distinguish between a subjective component of understanding (perceived understanding) and an objective component (actual understanding).

For perceived understanding, we measure how well participants felt that they understood an explanation of differential privacy. We adapt three questions from [56] for measuring perceived comprehension of a text and a fourth adapted from [26]. All four questions use 6-pt semantic scales.

For actual understanding, we measure how well participants actually understood an explanation by asking them concrete questions about differential privacy. Two questions (#1 and #2 in Table 5) are drawn from [87]. Two additional questions (#3 and #4 in Table 5) are related to the privacy parameter, since our explanations were developed with the specific goal of conveying information about the privacy parameter.

In the second survey, some participants only see low-transparency explanations that do not convey the necessary information to answer these questions. Therefore, one of the answer options was "I do not have enough information to answer this question." Prior work has shown that survey takers gravitate toward these choices [49], rather than thinking carefully about whether the relevant

[4]Uber actually uses central differential privacy rather than local differential privacy, but the selected excerpt applies equally well to either model.

information is present [48]. To deal with this issue, in the second experiment, participants who select the "not enough information" option were prompted to make their best guess; prior work has found this strategy to be useful [84]. Participants who view low-transparency explanations get the question correct if they select the "not enough information" option; their best guess does not factor into their comprehension score. Participants who view high-transparency explanations get the question correct if they select the correct answer on the first try or if they select "not enough information" but subsequently select the correct answer when prompted to make their best guess.

After the launch of the first survey, it became clear that participants found our comprehension questions relatively difficult, so three additional, easier questions were added for the second survey (#5-7 in Table 5).

3.4.2 Willingness To Share. The survey asks participants whether they would be willing to share their browsing history data before the privacy protection is described. This provides a baseline for participants' willingness to share and helps pinpoint the actual effect of the explanation of privacy protection; we would not want to infer that a particular explanation has persuaded a participant to share information when in fact this participant would have been perfectly happy to share the information without any privacy protection. Therefore, it is useful to ask about willingness to share twice, both before and after participants read the explanation of privacy protections.

Prior work has frequently asked about hypothetical willingness to share information rather than observing willingness to share directly [11, 87]. However, hypothetical willingness to share is an

Table 5. Comprehension Questions. All questions are multiple choice. Correct answers are in parentheses. For questions #3 and #4, the correct answer depends on the experimental condition. The full survey instrument is included in the supplementary materials.

1. If someone shares their information, will the researchers know with certainty which websites that person visited? *(No.)*
2. If someone shares their information with us, and an attacker steals the information, will the attacker know with certainty which websites that person visited? *(No.)*
3. If someone shares their information, the method will change how many of their true answers for the 200 websites of interest? *(One of the following: A few–for about 1 or 2 websites, their true answer will be changed. / Some–for about 50 websites, their true answer will be changed. / Many–for about 100 websites, their true answer will be changed. / I do not have enough information to answer this question.)*
4. If someone shares their information, how accurately will the researchers understand that person's individual internet usage from the information they share? *(One of the following: Very accurately [a pretty good guess] / Somewhat accurately [a bit better than a random guess] / Inaccurately [not much better than a random guess] / I do not have enough information to answer this question.)*
5. For those who choose to share their information, what kind of privacy protection will we use to protect that information? *(Local differential privacy)*
6. How does the method protect an individual's privacy? *(Changing some of their answers)*
7. How does the method help the researchers collecting the data? *(They can see meaningful information emerge from large datasets.)*

imperfect proxy for actual disclosure behavior; previous work has found both that individuals tend to underestimate their actual willingness to share sensitive information and that risk perceptions are more strongly related to behavioral intention (i.e. hypothetical willingness to share) than to actual disclosure behavior [58]. In other words, it is often the case that people do not actually act in accordance with their risk perceptions. As a result, it can be difficult to shift user behavior. Since we are most interested in actual behavior, we measure willingness to share by actually asking participants to share their browser histories with us. If participants agree to share this data, the study shows an animation of a "progress bar" that suggests the browser history data is being uploaded. However, no browser history information is collected, and participants are debriefed about this deception at the end of the study.

Of course, it does not make sense to ask participants to share their data twice. Therefore, the first ask is hypothetical and only in the second ask are participants actually asked to share their information.

3.4.3 Privacy Concerns. An individual's decision about whether or not to share data will be informed by the concerns that this individual holds regarding information privacy. We use the Internet Users' Information Privacy Concerns scale to measure privacy concern [52]. In addition to the ten questions for this scale, we add two additional related questions—adapted from previous work—that are relevant to our setting; we ask how frequently participants falsify personal information online [52, 62] and whether or not they ever try to hide their online activities from others [65]. The ordering of these two added questions is randomized. All questions use a 5-pt scale.

3.4.4 Sensitivity. Since information sensitivity can affect users' willingness to share data [87], participants are asked to rate the sensitivity of the browsing history information we request from them. With wording adapted from [70], participants rate the perceived sensitivity of the requested browsing history information. Participants also rate the harm that could result from this information being leaked. Both questions use a 5-pt scale and are averaged to produce the measure of sensitivity.

3.4.5 Browser Usage. When asked to share internet browsing data, people who regularly browse the Internet may behave differently than people who use Internet browsers less frequently. Therefore, participants are asked about their frequency of browser use.

3.5 Recruitment

For both surveys, English-speaking participants over the age of 18 were recruited through Qualtrics' paid recruiment service to be approximately representative of the United States population in terms of educational attainment, age, race/ethnicity, and gender. The surveys were also hosted on Qualtrics. Participants who failed the attention check questions—described in Section 3.6—were screened out of the survey. Some participants' responses were deleted for speeding (first survey: < 220 seconds, second survey: < 240 seconds), straightlining, or entering gibberish in the free response textbox. Finally, to achieve this representative sample, some participants were screened out after entering their demographic information. This resulted in 365 total participants for the first experiment and 308 for the second experiment. For these participants, the median completion time was approximately 7 minutes (7m 24s) for the first survey and 9 minutes (9m 7.5s) for the second survey. Participants who completed either survey were paid by Qualtrics. The first survey was conducted in March 2021 and the second in April 2021. The supplementary materials include the full participant demographics.

3.6 Experimental Protocol

Both experiments use a between subjects design. Participants begin the survey by answering demographic questions. Next, the survey explains the general concept of internet browsing histories, providing a brief description for participants with low technology literacy. Then participants are prompted to reflect for a moment on their browser usage and answer the sensitivity questions. It is by design that participants reflect on the sensitivity of this information before they are asked to share it; only at the end of the sensitivity questions do we ask for the first time whether the participant would be willing to share their information.

We then explain the differentially private mechanism, with participants assigned at random to one of the four experimental conditions for the first experiment and to one of the nine conditions for the second experiment. For the second experiment, a timing mechanism prevents participants from advancing to the next section until they have spent at least one minute reading through the explanation. Participants are also required to click several times in order to continue reading; this was intended to slow participants down and encourage them to read carefully.

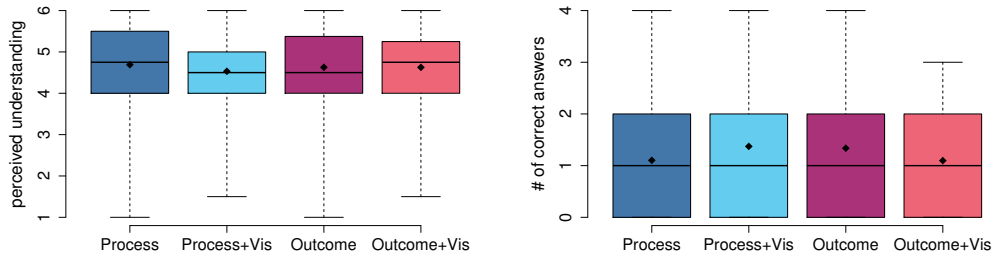
Next, four questions serve as an attention check in order to screen out participants who have not actually read the explanation. The questions are designed to be easy to answer, and the participants are able to reread the explanation in order to answer the questions; these choices aim to prevent screening out participants who have read the explanation carefully but may struggle with reading or have limited working memory capacity. Participants who fail to answer the first two questions correctly are given a second chance and presented with the remaining two questions. If these questions are not answered correctly, participants are screened out of the study.

Participants who pass this check proceed by completing items for the dependent measures: perceived understanding, actual understanding, and willingness to share. Participants can re-read the explanation as necessary while answering these questions. If participants report being willing to share their browsing history, an animation suggests this information is being uploaded automatically. Participants also answer an open-ended question about why they were or were not willing to share their data. Finally, participants complete questions related to the remaining covariates: general privacy attitudes and internet usage. Before they submit the survey, participants are given access to a debriefing document and informed that no browsing history data was actually collected. At the completion of the survey, resources are provided so that participants can learn about best privacy practices, such as clearing browsing histories.

4 ANALYSIS

4.1 Testing Hypotheses

To test our hypotheses for the first experiment, we used ANCOVA to compare the objective comprehension scores for the four treatment groups, controlling for education. Since only two participants declined to provide their educational background, these participants were dropped from this analysis. To test our hypotheses for the second experiment, we fit a logistic regression model that models willingness to share as a function of 1) the choice of explanation, 2) the choice of privacy parameter ε, 3) the interaction between these two variables, 4) perceived sensitivity, and 5) the privacy concerns measure. The privacy parameter was treated as an ordinal variable. Three participants who declined to answer questions for the covariates were dropped from this analysis. We fit one regression model on the full set of responses for the second survey and one regression model on a subset of responses, excluding participants who had been willing to share their data even before reading the description of differential privacy. The goal of this second regression was to test if our hypotheses would hold for these more privacy-conscious participants who were not willing to share their data without privacy protection.



(a) Perceived Understanding (b) Actual Understanding

Fig. 1. Perceived and actual understanding for Experiment 1. Participants were overconfident in their understanding. Most participants felt that they understood the explanations well (left), but answered no more than one question correctly (right). Performance was similar across all four conditions.

4.2 Comparing Comprehension

We compared participants' performance on two closely related comprehension questions (#1 and #2 in Table 5)—one asks whether the researchers will be uncertain about participants' true answers while the other asks the same question about an external attacker. The question about the external attacker is more aligned with how discussions about security and privacy are traditionally framed. The question about whether the researchers will know the true responses is trickier—this idea of protecting privacy from the same people with whom they are sharing data will be less familiar to participants. We compare performance on the two questions using the two proportion z-test; the responses from the two surveys are analyzed together.

4.3 Qualitative Analysis

To analyze the qualitative data from participants about why they chose (not) to share their information, two researchers used an iterative open coding approach [9]. The researchers performed two rounds of coding. First, both researchers familiarized themselves with the responses from the first survey. Next, the researchers agreed upon a set of themes. When the responses came in from the second survey, the researchers again began by familiarizing themselves with the data, and they decided that some additional themes should be added. Finally, all responses from both surveys were analyzed together and assigned themes; multiple themes could be assigned to a single response. Disagreements were remedied through discussion until agreement was reached. The codebook can be found in the supplementary materials.

5 RESULTS

The analysis of our results revealed no statistically significant effects. Interestingly, we found that most participants had made up their minds about whether or not to share their data before they even read about the privacy protection; therefore, these explanations had little effect on participants' behavior. Below we discuss the results in detail for both experiments and for the qualitative analysis. Further details are included in the supplementary materials.

5.1 Experiment 1: Developing Good Explanations of Differential Privacy

Participants were overconfident in their understanding of differential privacy; the scores for perceived understanding were substantially higher than their actual understanding (Figure 1). Participants may not have spent enough time reading the explanations carefully; many people are

used to scrolling through privacy policies and clicking accept without actually reading them [31, 55, 59]. Participants who took more than 7 minutes to complete the survey (roughly 50% of participants) answered 0.47 more questions correctly — over a 10 percentage point improvement — compared to those who finished more quickly. To address this issue, experiment 2 set a minimum time for reading the explanation.

As hypothesized, the outcome-focused explanation outperformed the process-focused explanation; the average percentage of comprehension questions answered correctly was greater for outcome (33%) than process (28%) focused explanations. However, the difference was not significant. Figure 1b shows the distribution of the number of questions answered correctly for each condition.

Adding a visual component did not uniformly improve participants' understanding. The process-focused visual may have been helpful for participants. It slightly improved the average percentage of comprehension questions answered correctly (35%), but the outcome-focused visual decreased this performance (27%). Again, the differences were not statistically significant.

As none of the differences were statistically significant (ANCOVA comparing all four explanations gave $p = 0.1$), we selected the outcome explanation for the second experiment. Of all four explanations, process+vis resulted in the highest average comprehension score; however, the differences were small, and a text-based explanation would serve the fairest comparison for the text-only industry explanations.

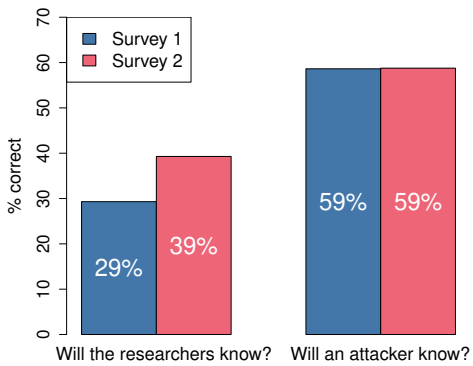
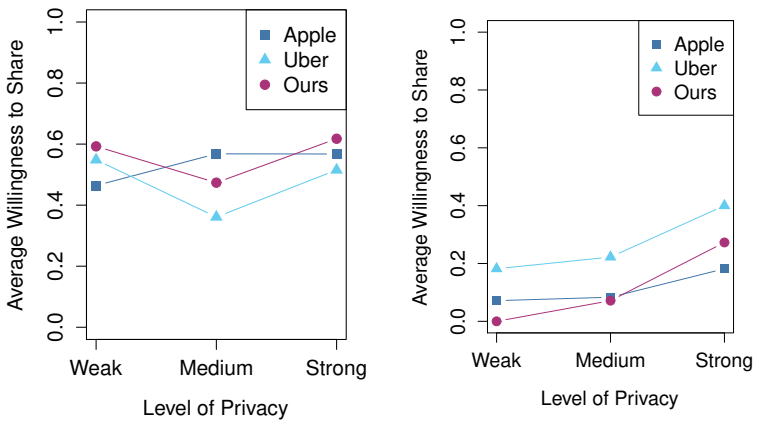


Fig. 2. Performance on two paired questions. Most participants did not understand that the researchers would be uncertain about which websites participants truly visited, though most participants did understand that they would have this protection from attackers; the difference in performance is statistically significant ($p < 10^{-19}$).

Two paired comprehension questions identified one challenge for explaining differential privacy (Figure 2). Most participants correctly identified that attackers who stole survey data from the researchers would be unable to determine any participant's true browsing history with certainty. However, most participants incorrectly believed that the researchers would have access to participants' true browsing histories. Participants are likely more familiar with privacy tools to protect from hackers than with tools that allow for sharing information while preserving privacy. It may be counterintuitive that the information they have agreed to share will be obfuscated even for the researchers they agreed to share it with.

There are several factors that may have caused the poor performance on the comprehension questions. The fact that perceived understanding was high suggests that the readability of the explanations was not the problem, although participants may have found some of the comprehension questions confusing. As discussed previously, a lack of attentive reading is likely one reason for the low comprehension. Most people are habituated to skimming through privacy policies or clicking "accept" without reading [59]. Therefore, even though our explanations were designed to be much more accessible than the typical privacy policy, people may have skimmed our explanations out of habit. Our results also indicate that most participants simply were not very concerned about their data privacy in this context—yet another reason they may not have paid close attention to the explanations. Another factor may be that certain aspects of differential privacy are simply counterintuitive. For example, participants may find it strange to think about hiding information



(a) All participants

(b) Initially unwilling to share

Fig. 3. Willingness to share for the three explanations and three settings of the privacy parameter. These figures show the averages for all participants (left) and participants who were initially unwilling to share their information (right). For participants who saw our explanation in the weak privacy setting, the only participants who agreed to share were those who were already willing to share before learning about differential privacy.

from the same people with whom they are sharing data—this would explain the poor performance on the question about whether the researchers would know participants' true information (Figure 2). Finally, two of the questions (#3 and #4 in Table 5) were expected to be more difficult, since they required participants to reason about probabilities.

5.2 Experiment 2: Measuring Privacy Theater

Contrary to our expectations, the choice of privacy parameter ε and choice of explanation had no significant effect on willingness to share ($p = 0.1 - 0.99$). Most participants made up their minds about whether or not to share their data before reading about the privacy protection.

The survey asks participants about their willingness to share twice. Early in the survey, participants are asked about their willingness to share their browsing histories. Later, after explaining the protection offered by differential privacy, participants are asked whether they agree to share this information via an automated upload process. Most participants (76%) do not change their minds after reading about the privacy protection (Figure 4).

The open-ended responses (see Section 5.3) provide some context for this finding. Many participants are willing to share their data because they trust researchers and want to help or because they do not consider browsing data particularly sensitive; for

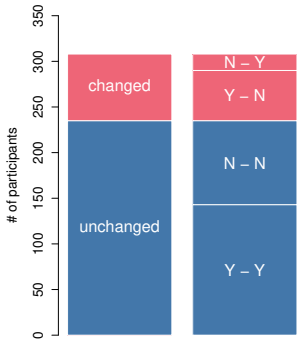
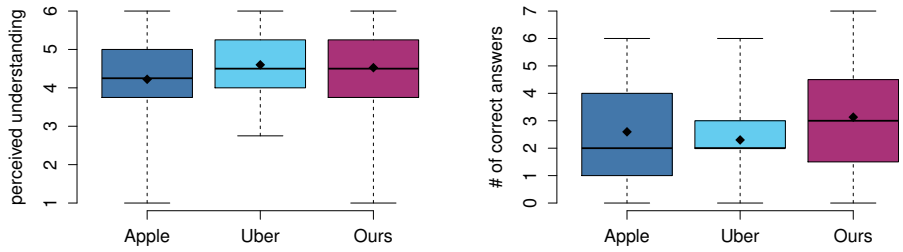


Fig. 4. Willingness to share before and after differential privacy explanation. Most participants' answers did not change (left) when asked about their willingness to share a second time. Of those who did, the differential privacy explanations were more successful at convincing users not to share (yes-to-no) than to share (no-to-yes) (right).



(a) Perceived Understanding (b) Actual Understanding

Fig. 5. Perceived and actual understanding for Experiment 2. As in Experiment 1, participants' perceived understanding (left) is greater than their actual understanding (right), though the added questions improved comprehension scores compared to Experiment 1. The Apple explanation gave precise values of ε without explaining the implications of these values, which may explain its lower perceived understanding scores.

these participants, the promises of privacy protection are irrelevant. The set of participants for whom the privacy protection really mattered was relatively small—only 36% of participants were unwilling to share their data when first asked, even before any privacy protection was mentioned. Of those initially unwilling to share, many had strong negative reactions to being asked to share this information or found it too sensitive to share. The set of participants for whom the privacy protection made the biggest difference were those who changed their minds after reading the differential privacy explanation. This group was quite small (n=73).

Although we did not find a statistically significant relationship between the variables of interest, we noticed an intriguing result. In the high-transparency, weak-privacy setting, no one agreed to share, unless they had already agreed to share before they read any description of privacy protection. This suggested that for privacy-conscious participants, H4 may actually hold. In other words, our explanation may be effective in helping privacy-conscious users make more appropriate data-sharing decisions. However, we did not have enough data to discern whether this pattern was meaningful, since so few participants were initially unwilling to share data (only nine for this condition). To investigate this pattern further, we conducted a larger replication study. While the overall rates of (un)willingness to share remained consistent, this study failed to provide evidence to support our hypotheses. The details about this study are included in the supplementary materials. Although our failure to find a significant effect of explanation on sharing behavior seems to contradict some prior work [87], a more recent paper found that descriptions of differential privacy had a significant effect on privacy expectations yet not on willingness to share. The authors suggest that effective explanations may need to be tailored to the concerns of individual users [43]. Our results, however, show that it may be difficult to shift people's preconceived feelings about what information should or should not be shared.

5.3 Qualitative Results

Both surveys asked participants to explain why they decided to share or not to share their information. This qualitative data helped clarify why some participants chose to share their information or not, as well as why some changed their minds after reading about the privacy protection offered by local differential privacy.

5.3.1 Reasons for Sharing and Withholding Information. Most participants did not change their willingness to share based on the explanation of privacy protections. This participant made it clear that the protection offered by local differential privacy was irrelevant to their decision: *No matter how it is shared, it does not seem like a good idea* [P588]. Other participants listed factors such as distrust of the researchers or the survey platform (n=37), fears of getting "scammed" or "hacked" (n=27), general discomfort (n=40), or sensitivity of the requested information (n=56). For example, two participants mentioned concerns about others seeing pornographic websites that they had visited.

Participants who were willing to share their information even before learning about the privacy protections and followed through in sharing listed a variety of reasons for doing so. Many participants—mainly in the first survey—mentioned concerns about being able to complete the survey and receive compensation (n=33), for example, writing that, *I need money badly so it's worth the weird risk* [P129] and *I really need the money from this survey to help take care of my family until pay day. I really hope its not a scam* [P192]. As a result, the second survey clarified that participants would be able to complete the survey and receive payment regardless of whether or not they agreed to share browsing data. These responses point to a larger issue; all too often, the choice of whether to share information is hardly a choice at all [23, 51].

Another common reason that participants agreed to share their information was that they trusted the researchers or the survey platform (n=46). In other words, these participants agreed to share their information because they trusted the data recipients. This finding is consistent with prior work that has shown that the reputation of the data recipient can be an important factor in data sharing decisions [4].

Other participants who were wiling to share their information listed reasons such as a desire to help the researchers (n=66), curiosity (n=31), and a sense that they had "nothing to hide" (n=82). For example, one participant wrote: *Research and study is important! I have mostly nothing to hide* [P125]. Another commonly expressed theme was indifference (n=27); when asked why they decided to share their data, many responded with another question—why not?

An interesting theme that arose in many of the responses was a concern about the scope of data collection more broadly (n=12). Some participants felt that so much data was being collected about them already that they might as well share more: *I decided to share because this information is already out of my hands given how my ISP has free access, might as well contribute to a study while i'm at it* [P30]. On the other hand, others expressed a desire to hang on to what they perceived as the last little bit of privacy that they had left:

> *I refuse to share this information because all of my information is already online, theres no need to go through my PRIVATE browsing history. My phone is my phone, and sorry but no one has a right to that. Privacy doesnt exist in this world anymore, the least I could do is keep my phone away from the world* [P229].

It is interesting that both participants who shared their information and participants who declined to share cited these same kinds of concerns when asked to justify their choices. While some participants felt the need to exert extra effort to protect their privacy in a world of ubiquitous data collection, others expressed feelings of resignation and fatalism [18].

5.3.2 Perceptions of the privacy protection. Although most participants did not mention the protection provided by differential privacy in their responses, some participants did feel comforted by this privacy protection and trusted it to protect their information (n=40): *It seems like the process you're using will protect my information* [P294]. However, some participants felt suspicious and did not trust that the mechanism would protect their data (n=11): *The process didn't seem safe. I'm not*

sure what information would be changed. This seems like a hoax [P332][5]. Other participants seemed concerned by the description of the "automatic upload" process. For example, one participant who initially expressed being willing to share information later declined to share, explaining that: *An upload of my information is a little daunting* [P199]. These participants were more concerned about a mysterious script running on their computer than about sharing browsing data. Finally, some participants remained confused after reading the explanation of the privacy protection (n=14). Nevertheless, most participants rated their own understanding of the explanations highly (Figure 1 and Figure 5).

6 LIMITATIONS

Our study has several limitations, some of which point toward directions for future work. The first limitation is that since most participants were perfectly willing to share their data, we can not draw firm conclusions about the behavior of individuals who were actually concerned about privacy. In order to better understand how people behave when concerned about their privacy, future work might try to explicitly target privacy-conscious individuals—for example, the topic of privacy could be emphasized in recruitment materials.

A second limitation concerns our qualitative results. Although the responses to the open-ended survey question provided valuable insight, most responses were quite short—a few sentences at most. A more thorough understanding of individuals' thought processes might be obtained through interviews or laboratory studies. These kinds of studies might also shed light on any misconceptions that participants may have about differential privacy.

A third limitation of this work is that behavior in a research study may differ significantly from behavior in everyday life. For example, we asked participants to read the description of differential privacy carefully. In everyday life, when asked to share their data, many people will not read the fine print. Furthermore, the participants knew that they were sharing their data with researchers. Some participants specifically mentioned this fact as a reason that they trusted us and were willing to share their data. People may behave differently when asked to share data with corporate (rather than academic) entities. Future work might study situations in which people encounter differential privacy "in the wild."

7 DISCUSSION

7.1 False Choices

Although we did not intend to force participants to choose between their privacy and their paycheck, some participants did feel compelled to make this choice due to a lack of clarity in the first survey. Several participants explained that despite their fears, they felt that they had no real choice, since they desperately needed the money. These responses offer a glimpse into the myriad of ways that marginalized people in particular are asked to surrender their privacy. Prior work has examined situations in which people living in poverty are forced to sacrifice privacy in order to access resources [23]. Studies that limit their focus to "*the often-studied White, American, middle-class subject*" will miss important aspects of the way that privacy and surveillance operate [53].

7.2 Digital Resignation

Several participants expressed concerns about the ubiquity of digital surveillance. Some of these participants opted to share their data, because they felt that trying to protect their privacy was a losing battle. Prior work has observed similar phenomena, sometimes referred to as "*privacy cynicism*", "*surveillance realism*", or "*digital resignation*" [15, 18, 35]. In fact, feelings of helplessness

[5]This participant was assigned to the outcome+vis condition.

regarding surveillance appear to be widespread; a recent survey found that more than 60% of Americans agree that "it is not possible to go through daily life without being tracked" by companies or by the government [7].

Awareness of digital resignation should inform the design of privacy-enhancing technologies; for example, tools that increase awareness of online tracking without actually offering improved control over one's data may unintentionally lead to increased resignation rather than increased engagement with privacy-related issues [12, 79]. More work is needed to better understand the causes of digital resignation and to propose solutions for overcoming these feelings of futility, inevitability, and hopelessness. There may be particular strategies that help people regain a sense of agency. For example, in the context of environmental activism, Sarah Jaquette Ray argues that "*the perception that social change happens only on an individual scale creates defeatism*" and that "*if we see ourselves working collectively rather than individually, we can rest assured that we are contributing to a larger web of movement*" [66]. Perhaps collective strategies for resisting surveillance (e.g. obfuscation on Instagram [57]) and PETs that support such strategies (e.g. AdNauseam [37]) could be an effective way to combat digital resignation.

7.3 Making Sense of Privacy Promises

The results of our experiments suggest that for many people, factors such as the purpose of data collection or a (dis)trust of data collectors are more important than promises of privacy protection when deciding whether or not to share information. If people trust the data collector and if people expect to share in the benefits of data collection, they may feel that differential privacy is unnecessary. Perhaps discussions about optimal privacy-utility tradeoffs have at times overshadowed equally important discussions about the distribution of utility—in other words, discussions about who benefits from data analysis.

Even though explanations of differential privacy seem to only matter to a small set of participants, for companies with millions of users, even small effects matter. Future work could help clarify both 1) when users care about promises to protect privacy and 2) how users respond to such promises. As more tech companies begin to tout their privacy protection features in advertising campaigns, these questions will become increasingly important [61].

Our results revealed that some aspects of differential privacy are counterintuitive to users and may be difficult to convey even in well-crafted explanations. For example, users do not expect to need to protect themselves from the people they are sharing their data with. In practice, industry explanations may not be well-crafted; in fact, they may be intentionally difficult to understand [18]. And even within a single industry (e.g., web browsers) there are a wide variety of technical approaches. As a result, people searching for useful information about ways to protect their privacy may grow frustrated with the "*widespread obfuscatory communication practices used by companies across the digital-media landscape*" [18] and struggle to understand which privacy promises are actually meaningful [40]. In trying to make sense of this landscape, people turn to journalists, to family, to friends, and to other trusted sources [64, 67]. Understanding how people engage within communities to learn and make sense of this information could help establish important avenues for intervention. And given the special opportunity that journalists have to help readers make sense of information about privacy that often lies buried in unintelligible privacy policies [24], the way that journalism shapes privacy practices is also worthy of further study.

More broadly, many questions remain about the rhetorical functions of privacy-enhancing technologies. Companies have a variety of avenues for communicating promises of privacy to their users—through privacy policies, app notifications, press releases, blog posts, and more. Understanding how this messaging shapes individual perceptions of technology companies is important. But it is worth asking not only how PETs and their associated messaging affect users' interactions

with particular apps or companies but also how they shape larger conversations around online privacy [33, 36, 37, 69]. The term "*privacy*" does not have a single, straightforward agreed-upon meaning [5, 74]. Thus, in addition to whatever technical affordances they provide, PETs may make implicit claims about the very meaning and value of privacy. For example, AdNauseam—a browser extension that hides ads while employing obfuscation mechanisms to sabotage advertising surveillance—serves both a practical, protective purpose and an expressive purpose; AdNauseam actively promotes the conception of privacy as a collective good [37].

7.4 Other Approaches for Addressing Privacy Theater

The problem of privacy theater is not unique to differential privacy [13, 14, 34, 73]. As a consequence, addressing the problem is likely to require an array of partial solutions. Carefully developed explanations can be one part, and prior work has discussed other approaches [20, 39]; still, more work is needed. While it is worthwhile to develop accessible explanations of privacy technologies, many people who are accustomed to the opaque, jargon-laden style of privacy policies may simply skim or skip such explanations out of habit. Furthermore, the burden of ensuring that the privacy protection offered by a particular company is adequate should not rest with individuals. Many researchers have noted that by offering control to users, technology firms can shift burdens from themselves to individual users [3, 77, 81]. Instead, given the power asymmetries that exist, policy makers and others should reflect on societal goals around privacy protection. In addition, there may be technical approaches like expert audits and other accountability mechanisms that can help drive organizational responses [20, 87].

8 CONCLUSION

In this work, we explore how people decided whether or not to share their data when offered the protection of differential privacy. We develop explanations of differential privacy for non-experts, but find that certain aspects of differential privacy remain challenging for people to understand. We also test whether explanations that offer little transparency into the role of the privacy parameter can be misused to trick people into sharing more data than they otherwise would. However, we find that most people are perfectly willing to share their browsing data, regardless of whether they are offered the protection of differential privacy; other factors seem to dominate decision making around the sharing of browsing data. We point to the need for future work in order to better understand the effects of differential privacy and other privacy-enhancing technologies.

ACKNOWLEDGMENTS

MAS was supported by a Qualcomm Fellowship. We also thank the anonymous reviewers for their helpful comments, suggestions, and insights.

REFERENCES

[1] John M. Abowd. 2018. The U.S. Census Bureau Adopts Differential Privacy. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, London United Kingdom, 2867–2867. https://doi.org/10.1145/3219819.3226070

[2] Hazim Almuhimedi, Florian Schaub, Norman Sadeh, Idris Adjerid, Alessandro Acquisti, Joshua Gluck, Lorrie Faith Cranor, and Yuvraj Agarwal. 2015. Your Location has been Shared 5,398 Times! A Field Study on Mobile App Privacy Nudging. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. Association for Computing Machinery, Seoul, Republic of Korea, 787–796. https://doi.org/10.1145/2702123.2702210

[3] Mike Ananny and Kate Crawford. 2018. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *new media & society* 20, 3 (2018), 973–989.

[4] Eduardo B Andrade, Velitchka Kaltcheva, and Barton Weitz. 2002. Self-disclosure on the web: The impact of privacy policy, reward, and company reputation. *ACR North American Advances* (2002).

[5] Payal Arora. 2019. Decolonizing privacy studies. *Television & New Media* 20, 4 (2019), 366–378.

[6] Differential Privacy Team at Apple. 2017. Learning with privacy at scale. *Apple Machine Learning Journal* 1, 8 (2017).

[7] Brooke Auxier, Lee Rainie, Monica Anderson, Andrew Perrin, Madhu Kumar, and Erica Turner. 2019. Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information. Pew Research Center. https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/

[8] Wei Bai, Moses Namara, Yichen Qian, Patrick Gage Kelley, Michelle L. Mazurek, and Doowon Kim. 2016. An Inconvenient Trust: User Attitudes toward Security and Usability Tradeoffs for Key-Directory Encryption Systems. 113–130. https://www.usenix.org/conference/soups2016/technical-sessions/presentation/bai

[9] Richard E Boyatzis. 1998. *Transforming qualitative information: Thematic analysis and code development.*

[10] Finn Brunton and Helen Nissenbaum. 2015. *Obfuscation: A user's guide for privacy and protest.* MIT Press.

[11] Brooke Bullek, Stephanie Garboski, Darakhshan J. Mir, and Evan M. Peck. 2017. Towards Understanding Differential Privacy: When Do People Trust Randomized Response Technique?. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, Denver, Colorado, USA, 3833–3837. https://doi.org/10.1145/3025453.3025698

[12] Jessica Colnago, Yuanyuan Feng, Tharangini Palanivel, Sarah Pearman, Megan Ung, Alessandro Acquisti, Lorrie Faith Cranor, and Norman Sadeh. 2020. Informing the design of a personalized privacy assistant for the internet of things. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[13] Federal Trade Commission. 2014. Snapchat settles FTC charges that promises of disappearing messages were false.

[14] Federal Trade Commission. 2019. FTC imposes $5 billion penalty and sweeping new privacy restrictions on Facebook.

[15] Lina Dencik and Jonathan Cable. 2017. The advent of surveillance realism: Public opinion and activist responses to the Snowden leaks. *International Journal of Communication* 11 (2017), 763–781.

[16] Verena Distler, Carine Lallemand, and Vincent Koenig. 2020. Making Encryption Feel Secure: Investigating how Descriptions of Encryption Impact Perceived Security. In *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS PW)*. 220–229. https://doi.org/10.1109/EuroSPW51379.2020.00037

[17] Paul Dourish, Rebecca E Grinter, Jessica Delgado De La Flor, and Melissa Joseph. 2004. Security in the wild: user strategies for managing security as an everyday, practical problem. *Personal and Ubiquitous Computing* 8, 6 (2004), 391–401.

[18] Nora A Draper and Joseph Turow. 2019. The corporate cultivation of digital resignation. *New Media & Society* 21, 8 (2019), 1824–1839.

[19] Cynthia Dwork. 2006. Differential Privacy. In *Proceedings of the 33rd International Conference on Automata, Languages and Programming - Volume Part II (ICALP'06)*. Springer-Verlag, Berlin, Heidelberg, 1–12. https://doi.org/10.1007/11787006_1

[20] Cynthia Dwork, Nitin Kohli, and Deirdre Mulligan. 2019. Differential Privacy in Practice: Expose your Epsilons! *Journal of Privacy and Confidentiality* 9, 2 (Oct. 2019). https://doi.org/10.29012/jpc.689

[21] Nico Ebert, Kurt Alexander Ackermann, and Peter Heinrich. 2020. Does Context in Privacy Communication Really Matter? — A Survey on Consumer Concerns and Preferences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, Honolulu, HI, USA, 1–11. https://doi.org/10.1145/3313831.3376575

[22] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. 2014. RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response. *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security - CCS '14* (2014), 1054–1067. https://doi.org/10.1145/2660267.2660348 arXiv: 1407.6981.

[23] Virginia Eubanks. 2018. *Automating inequality: How high-tech tools profile, police, and punish the poor.* St. Martin's Press.

[24] Benjamin Fabian, Tatiana Ermakova, and Tino Lentz. 2017. Large-Scale Readability Analysis of Privacy Policies. In *Proceedings of the International Conference on Web Intelligence* (Leipzig, Germany) *(WI '17)*. Association for Computing Machinery, New York, NY, USA, 18–25. https://doi.org/10.1145/3106426.3106427

[25] Chris Fennell and Rick Wash. 2019. Do Stories Help People Adopt Two-factor Authentication? (2019), 5.

[26] Philip M. Fernbach, Todd Rogers, Craig R. Fox, and Steven A. Sloman. 2013. Political Extremism Is Supported by an Illusion of Understanding. *Psychological Science* 24, 6 (June 2013), 939–946. https://doi.org/10.1177/0956797612464058

[27] A. Ferreira and G. Lenzini. 2015. An analysis of social engineering principles in effective phishing. In *2015 Workshop on Socio-Technical Aspects in Security and Trust*. 9–16. https://doi.org/10.1109/STAST.2015.10

[28] Dorothy S Fidler and Richard E Kleinknecht. 1977. Randomized response versus direct questioning: Two data-collection methods for sensitive information. *Psychological Bulletin* 84, 5 (1977), 1045.

[29] Patricia Garcia, Tonia Sutherland, Marika Cifor, Anita Say Chan, Lauren Klein, Catherine D'Ignazio, and Niloufar Salehi. 2020. No: Critical Refusal as Feminist Data Practice. In *Conference Companion Publication of the 2020 on Computer Supported Cooperative Work and Social Computing*. 199–202.

[30] C. S. Gates, J. Chen, N. Li, and R. W. Proctor. 2014. Effective Risk Communication for Android Apps. *IEEE Transactions on Dependable and Secure Computing* 11, 3 (May 2014), 252–265. https://doi.org/10.1109/TDSC.2013.58

[31] Joshua Gluck, Florian Schaub, Amy Friedman, Hana Habib, Norman Sadeh, Lorrie Faith Cranor, and Yuvraj Agarwal. 2016. How Short Is Too Short? Implications of Length and Framing on the Effectiveness of Privacy Notices. 321–340. https://www.usenix.org/conference/soups2016/technical-sessions/presentation/gluck

[32] Michael S Goodstadt and Valerie Gruson. 1975. The randomized response technique: A test on drug use. *J. Amer. Statist. Assoc.* 70, 352 (1975), 814–818.

[33] Seda Gürses, Arun Kundnani, and Joris Van Hoboken. 2016. Crypto and empire: the contradictions of counter-surveillance advocacy. *Media, Culture & Society* 38, 4 (2016), 576–590.

[34] Kashmir Hill. 2021. 'Do Not Track,' the Privacy Tool Used by Millions of People, Doesn't Do Anything. *Gizmodo* (2021). https://gizmodo.com/do-not-track-the-privacy-tool-used-by-millions-of-peop-1828868324

[35] Christian Pieter Hoffmann, Christoph Lutz, and Giulia Ranzini. 2016. Privacy cynicism: A new approach to the privacy paradox. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* 10, 4 (2016).

[36] Daniel C Howe. 2015. Surveillance countermeasures: Expressive privacy via obfuscation. *Datafied Research* 4, 1 (2015), 88–98.

[37] Daniel C Howe and Helen Nissenbaum. 2017. Engineering Privacy and Protest: A Case Study of AdNauseam.. In *IWPE@ SP*. 57–64.

[38] Harald Ibrekk and M Granger Morgan. 1987. Graphical communication of uncertain quantities to nontechnical people. *Risk analysis* 7, 4 (1987), 519–529.

[39] Matthew Jagielski, Jonathan Ullman, and Alina Oprea. 2020. Auditing Differentially Private Machine Learning: How Private is Private SGD?. In *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin (Eds.), Vol. 33. Curran Associates, Inc., 22205–22216. https://proceedings.neurips.cc/paper/2020/file/fc4ddc15f9f4b4b06ef7844d6bb53abf-Paper.pdf

[40] M. Janic, J. P. Wijbenga, and T. Veugen. 2013. Transparency Enhancing Tools (TETs): An Overview. In *2013 Third Workshop on Socio-Technical Aspects in Security and Trust*. 18–25. https://doi.org/10.1109/STAST.2013.11

[41] Bargav Jayaraman and David Evans. 2019. Evaluating differentially private machine learning in practice. In *28th USENIX Security Symposium (USENIX Security 19)*. 1895–1912.

[42] Verónica Juárez Ramos. 2018. *Analyzing the role of cognitive biases in the decision-making process*. IGI Global.

[43] Gabriel Kaptchuk, Rachel Cummings, and Elissa M Redmiles. 2021. "I need a better description": An Investigation Into User Expectations For Differential Privacy. In *2021 Workshop on Theory and Practice of Differential Privacy*.

[44] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W. Reeder. 2009. A "nutrition label" for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security (SOUPS '09)*. Association for Computing Machinery, Mountain View, California, USA, 1–12. https://doi.org/10.1145/1572532.1572538

[45] Patrick Gage Kelley, Lucian Cesca, Joanna Bresee, and Lorrie Faith Cranor. 2010. Standardizing privacy notices: an online study of the nutrition label approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. Association for Computing Machinery, New York, NY, USA, 1573–1582. https://doi.org/10.1145/1753326.1753561

[46] Rohit Khare. 2009. Privacy Theater: Why Social Networks Only Pretend to Protect You. *TechCrunch* (2009).

[47] J Peter Kincaid, Robert P Fishburne Jr, Richard L Rogers, and Brad S Chissom. 1975. *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel*. Technical Report. Naval Technical Training Command Millington TN Research Branch.

[48] Jon A Krosnick. 1991. Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied cognitive psychology* 5, 3 (1991), 213–236.

[49] Jon A Krosnick. 2018. Questionnaire design. In *The Palgrave handbook of survey research*. Springer, 439–455.

[50] Brian Y Lim, Qian Yang, Ashraf M Abdul, and Danding Wang. 2019. Why these Explanations? Selecting Intelligibility Types for Explanation Goals.. In *IUI Workshops*.

[51] Mary Madden, Michele Gilman, Karen Levy, and Alice Marwick. 2017. Privacy, poverty, and big data: A matrix of vulnerabilities for poor Americans. *Wash. UL Rev.* 95 (2017), 53.

[52] Naresh K. Malhotra, Sung S. Kim, and James Agarwal. 2004. Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. *Information Systems Research* 15, 4 (Dec. 2004), 336–355. https://doi.org/10.1287/isre.1040.0032 Publisher: INFORMS.

[53] Alice E Marwick and danah boyd. 2018. Privacy at the margins| understanding privacy at the margins—introduction. *International Journal of Communication* 12 (2018), 9.

[54] Alessandra Mazzia, Kristen LeFevre, and Eytan Adar. 2012. The PViz Comprehension Tool for Social Network Privacy Settings. In *Proceedings of the Eighth Symposium on Usable Privacy and Security* (Washington, D.C.) *(SOUPS '12)*. Association for Computing Machinery, New York, NY, USA, Article 13, 12 pages. https://doi.org/10.1145/2335356.2335374

[55] Aleecia M McDonald and Lorrie Faith Cranor. 2008. The cost of reading privacy policies. *Isjlp* 4 (2008), 543. Publisher: HeinOnline.

[56] David B. Miele and Daniel C. Molden. 2010. Naive theories of intelligence and the role of processing fluency in perceived comprehension. *Journal of Experimental Psychology: General* 139, 3 (2010), 535–557. https://doi.org/10.1037/a0019745

[57] Alfred Ng. 2021. Teens have figured out how to mess with Instagram's tracking algorithm. https://www.cnet.com/news/teens-have-figured-out-how-to-mess-with-instagrams-tracking-algorithm/

[58] Patricia A. Norberg, Daniel R. Horne, and David A. Horne. 2007. The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors. *Journal of Consumer Affairs* 41, 1 (2007), 100–126. https://doi.org/10.1111/j.1745-6606.2006.00070.x _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1745-6606.2006.00070.x.

[59] Jonathan A Obar and Anne Oeldorf-Hirsch. 2020. The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services. *Information, Communication & Society* 23, 1 (2020), 128–147.

[60] Lace Padilla, Matthew Kay, and Jessica Hullman. 2021. *Uncertainty Visualization.* 1–18. https://doi.org/10.1002/9781118445112.stat08296

[61] Mike Peterson. 2021. New iPhone privacy ad takes shots at other smartphones oversharing information. https://appleinsider.com/articles/20/09/03/new-iphone-privacy-ad-takes-shots-at-other-smartphones-oversharing-information

[62] Pew Internet Project. 2000. Trust and Privacy Online. https://www.pewresearch.org/internet/2000/08/20/trust-and-privacy-online/

[63] Irene Pollach. 2005. A typology of communicative strategies in online privacy policies: Ethics, power and informed consent. *Journal of Business Ethics* 62, 3 (2005), 221–235.

[64] Emilee Rader, Rick Wash, and Brandon Brooks. 2012. Stories as Informal Lessons about Security. In *Proceedings of the Eighth Symposium on Usable Privacy and Security* (Washington, D.C.) *(SOUPS '12)*. Association for Computing Machinery, New York, NY, USA, Article 6, 17 pages. https://doi.org/10.1145/2335356.2335364

[65] Lee Rainie, Sara Kiesler, Ruogu Kang, Mary Madden, Maeve Duggan, Stephanie Brown, and Laura Dabbish. 2013. Anonymity, privacy, and security online. *Pew Research Center* 5 (2013).

[66] Sarah Jaquette Ray. 2020. Claim Your Calling and Scale Your Action. In *A Field Guide to Climate Anxiety*. University of California Press, 52–79.

[67] Elissa M. Redmiles, Sean Kross, and Michelle L. Mazurek. 2016. How I Learned to Be Secure: A Census-Representative Survey of Security Advice Sources and Behavior. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (Vienna, Austria) *(CCS '16)*. Association for Computing Machinery, New York, NY, USA, 666–677. https://doi.org/10.1145/2976749.2978307

[68] Phillip Rogaway. 2015. The Moral Character of Cryptographic Work. *IACR Cryptol. ePrint Arch.* 2015 (2015), 1162.

[69] Jayshree Sarathy. 2022. From algorithmic to institutional logics: the politics of differential privacy. *Available at SSRN* (2022).

[70] Eva-Maria Schomakers, Chantal Lidynia, Dirk Müllmann, and Martina Ziefle. 2019. Internet users' perceptions of information sensitivity – insights from Germany. *International Journal of Information Management* 46 (2019), 142–150. https://doi.org/10.1016/j.ijinfomgt.2018.11.018

[71] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. 2017. Grad-CAM: Visual Explanations From Deep Networks via Gradient-Based Localization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.

[72] Karen L Soeken and George B Macready. 1982. Respondents' perceived protection when using randomized response. *Psychological bulletin* 92, 2 (1982), 487.

[73] Christopher Soghoian. 2011. An end to privacy theater: Exposing and discouraging corporate disclosure of user data to the government. *Minn. JL Sci. & Tech.* 12 (2011), 191.

[74] Daniel J Solove. 2008. Understanding privacy. (2008).

[75] Joshua Tan, Khanh Nguyen, Michael Theodorides, Heidi Negrón-Arroyo, Christopher Thompson, Serge Egelman, and David Wagner. 2014. The effect of developer-specified explanations for permission requests on smartphone user behavior. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*. ACM Press, Toronto, Ontario, Canada, 91–100. https://doi.org/10.1145/2556288.2557400

[76] Jun Tang, Aleksandra Korolova, Xiaolong Bai, Xueqiang Wang, and Xiaofeng Wang. 2017. Privacy Loss in Apple's Implementation of Differential Privacy on MacOS 10.12. *arXiv:1709.02753 [cs]* (Sept. 2017). http://arxiv.org/abs/1709.02753 arXiv: 1709.02753.

[77] Herman T Tavani and James H Moor. 2001. Privacy protection, control of information, and privacy-enhancing technologies. *ACM Sigcas Computers and Society* 31, 1 (2001), 6–11.

[78] Joseph Turow, Lauren Feldman, and Kimberly Meltzer. 2005. Open to exploitation: America's shoppers online and offline. *Departmental Papers (ASC)* (2005), 35.

[79] Joseph Turow, Michael Hennessy, and Nora Draper. 2015. The tradeoff fallacy: How marketers are misrepresenting American consumers and opening them up to exploitation. *Available at SSRN 2820060* (2015).

[80] Joseph Turow, Michael Hennessy, and Nora Draper. 2018. Persistent misperceptions: Americans' misplaced confidence in privacy policies, 2003–2015. *Journal of Broadcasting & Electronic Media* 62, 3 (2018), 461–478.

[81] Kristen Vaccaro, Christian Sandvig, and Karrie Karahalios. 2020. "At the End of the Day Facebook Does What It Wants" How Users Experience Contesting Algorithmic Content Moderation. *Proc. CSCW* (2020).

[82] Max Van Kleek, Reuben Binns, Jun Zhao, Adam Slack, Sauyon Lee, Dean Ottewell, and Nigel Shadbolt. 2018. X-ray refine: Supporting the exploration and refinement of information exposure resulting from smartphone apps. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[83] Nicholas Vincent, Hanlin Li, Nicole Tilly, Stevie Chancellor, and Brent Hecht. 2021. Data Leverage: A Framework for Empowering the Public in its Relationship with Technology Companies. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 215–227.

[84] Penny S Visser, Jon A Krosnick, Jesse Marquette, and Michael Curtin. 2000. Improving election forecasting: Allocation of undecided respondents, identification of likely voters, and response order effects. *Election polls, the news media, and democracy. New York: Chatham House* (2000).

[85] Stanley L. Warner. 1965. Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *J. Amer. Statist. Assoc.* 60, 309 (March 1965), 63–69. https://doi.org/10.1080/01621459.1965.10480775

[86] Ben Weinshel, Miranda Wei, Mainack Mondal, Euirim Choi, Shawn Shan, Claire Dolin, Michelle L Mazurek, and Blase Ur. 2019. Oh, the Places You've Been! User Reactions to Longitudinal Transparency About Third-Party Web Tracking and Inferencing. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*. 149–166.

[87] Aiping Xiong, Tianhao Wang, Ninghui Li, and Somesh Jha. 2020. Towards Effective Differential Privacy Communication for Users' Data Sharing Decision and Comprehension. In *2020 IEEE Symposium on Security and Privacy (SP)*. 392–410. https://doi.org/10.1109/SP40000.2020.00088

[88] Shoshana Zuboff. 2018. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (1st ed.).

Received April 2021; revised November 2021; accepted March 2022