"I need a better description": An Investigation Into User Expectations For Differential Privacy

Rachel Cummings
Columbia University
New York, United States
rac2239@columbia.edu

Gabriel Kaptchuk
Boston University
Boston, United States
kaptchuk@bu.edu

Elissa M. Redmiles
Max Planck Institute
for Software Systems
Saarbrücken, Germany
eredmiles@mpi-sws.org

ABSTRACT

Despite recent widespread deployment of differential privacy, relatively little is known about what users think of differential privacy. In this work, we seek to explore users' privacy expectations related to differential privacy. Specifically, we investigate (1) whether users care about the protections afforded by differential privacy, and (2) whether they are therefore more willing to share their data with differentially private systems. Further, we attempt to understand (3) users' privacy expectations of the differentially private systems they may encounter in practice and (4) their willingness to share data in such systems. To answer these questions, we use a series of rigorously conducted surveys ($n = 2424$).

We find that users care about the kinds of information leaks against which differential privacy protects and are more willing to share their private information when the risks of these leaks are less likely to happen. Additionally, we find that the ways in which differential privacy is described in-the-wild haphazardly set users' privacy expectations, which can be misleading depending on the deployment. We synthesize our results into a framework for understanding a user's willingness to share information with differentially private systems, which takes into account the interaction between the user's prior privacy concerns and how differential privacy is described.

CCS CONCEPTS

• **Security and privacy → Usability in security and privacy**.

KEYWORDS

Differential Privacy; Human Factors; Usable Security; Usable Privacy

ACM Reference Format:
Rachel Cummings, Gabriel Kaptchuk, and Elissa M. Redmiles. 2021. "I need a better description": An Investigation Into User Expectations For Differential Privacy. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security (CCS '21), November 15–19, 2021, Virtual Event, Republic of Korea.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3460120.3485252

CCS '21, November 15–19, 2021, Virtual Event, Republic of Korea

ACM ISBN 978-1-4503-8454-4/21/11…$15.00
https://doi.org/10.1145/3460120.3485252

1 INTRODUCTION

Differential privacy (DP) is a mathematically rigorous definition of privacy that has gained popularity since its formalization in 2006 [19]. DP facilitates the computation of aggregate statistics about a dataset while placing a formal bound on the amount of information that these statistics can disclose about individual data points within the dataset. Guaranteeing DP generally requires injecting carefully calibrated noise that hides individual datapoints while preserving aggregate level insights.

DP has become a leading technique used to meet the increasing consumer demand for digital privacy [4]. In the last few years, several companies have deployed DP. For instance, Apple uses DP to gather aggregate statistics on Emoji usage, which it uses to order Emojis for users [3, 67]. Uber uses DP to prevent data analysts within the company from stalking customers [36, 50], and Google uses DP to crowd-source statistics from Google Chrome crash reports [24].

The U.S. government has also begun to use DP. The United States Census Bureau is using DP to prevent information disclosure in the summary statistics it releases for the 2020 Decennial Census [1]. The use of DP in the Census means that nearly every person in the United States will have private data protected by DP.

Following in the footsteps of these earlier adopters, more companies have already announced their intentions to integrate differentially private techniques into their systems, e.g., [33, 49]. As a result, DP is becoming an increasingly consumer-relevant technology. Yet, little is known about whether end users value the protections offered by DP.

While DP is mathematically elegant and computationally efficient, it can be difficult to understand. Not only is DP typically defined mathematically, the privacy protections provided by DP are not absolute and require contextualization [17]. DP does not provide binary privacy (i.e., private or not private), but instead provides a statistical privacy controlled by unitless system parameters that are difficult to interpret (*i.e.*, the parameters ϵ and δ control the maximum amount of information that can leak about any individual entry in the dataset) [19]. Additionally, DP can be deployed in different security models, and the choice of model has significant impact on the types of adversarial behavior the system can tolerate. In the *local* model, users randomly perturb their information (with the help of the collection mechanism, e.g., their device) before sending it to a central entity in charge of analysis, called the curator [39]. In the *central* model users share their sensitive information directly, and the curator is trusted to perturb results that are released [21].

Differential Privacy from the user's perspective. The existing DP literature focuses on techniques for achieving DP [18–21, 27, 39, 46], with a small but growing body of work on legal and ethical implications of DP [11, 14, 53, 54]. Notably absent, however, is the voice of the end user, whose information may eventually be protected by DP and may benefit from its deployment [7]. Do users care about the information disclosures against which DP protects? Do users understand how DP protects them, and if so, do those protections influence their comfort with sharing information? As differentially private systems proliferate, it is increasingly important to answer such questions and understand DP from the *user's* perspective.

While a limited body of prior work has sought to understand how training users to understand DP influences willingness to share [9, 15, 69], we aim to answer broader questions regarding: (i) whether DP meets users' existing privacy needs, (ii) what expectations a potential user might have of a differentially private system, and (iii) whether existing, in-the-wild, descriptions of DP accurately set user expectations.

As a privacy-enhancing technology, DP is designed to prevent the unwanted information disclosure of user information to certain entities. However, it is not clear that these protections are meaningful to potential users. Additionally, it is not clear if DP provides the level of protection that potential users might hope.

Thus, in this work, we ask the following research questions:

(RQ1) Do potential users care about protecting their information against disclosure to the entities against which differentially private systems can protect?

(RQ2) Are potential users more willing to share their information when they have increased confidence that such information disclosures will not occur?

(RQ3) Do potential users expect differentially private systems to protect their information against disclosure? How does the way in which differential privacy is described impact their expectations?

(RQ4) Are potential users more willing to share their information when their information will be protected with differential privacy? How does the way in which differential privacy is described impact sharing?

We conduct two surveys with a total of 2,424 respondents to answer our research questions. We use vignette-based surveys to elicit respondents intended behavior, as such surveys have been found to well-approximate real-world behavior [26].

To address RQ1 and RQ2, we present each respondent with one of two information-sharing scenarios (sharing information with a salary transparency initiative or sharing medical records with a research initiative) and query respondents' privacy concerns. We then set their privacy expectations for those concerns (e.g., how likely information is to be leaked to a particular entity) and query if they would be willing to share their private information. Using the results of this survey, we examine how respondents' privacy concerns align with the protections offered by DP.

To address RQ3 and RQ4, we again present each respondent with one of the two information-sharing scenarios. We additionally tell respondents that their information will be protected by DP, as described by one of six descriptions.[1] We then query respondents' privacy expectations for these scenarios (e.g., whether their information could leak to various entities) and whether they would be willing to share their information. Using the results of this survey, we interrogate how accurately and effectively existing descriptions of DP set user expectations.

There is no "standard" deployment of DP, nor is there a "normal" way to describe its guarantees. In order to construct representative descriptions of DP to present to our participants, we systematically collected over 70 descriptions of DP written by companies, government agencies, news outlets, and academic publications. Through affinity diagramming qualitative analysis [8], we identify six main themes present in these descriptions, compose a representative description for each theme, and showed these representative definitions to respondents.

By describing DP as a potential user would encounter it in-the-wild, we gain a better understanding of how potential users are likely to respond to DP in practice. The nuances innate in DP make it easy for a prospective user to misunderstand what they are being promised. As such, a user seeking to choose the right privacy preserving system may find it difficult to make an informed choice. Getting this wrong can have real-world consequences: DP may be insufficient to protect a user's information against the types of threats about which they are concerned.

Summary of Findings. We find that users care about the kinds of information disclosure against which DP can protect (RQ1) and are more willing to share their private information when the risk of information disclosure to certain entities, specifically those for which disclosure would represent an inappropriate information flow [51], is not possible (RQ2).

Further, we find that descriptions of DP raise respondent's concrete privacy expectations around information disclosure (RQ3). This effect, however, varies by how DP is described: different descriptions of DP raise expectations for different kinds of information disclosure. These expectations, in turn, raise respondent's willingness to share information. However, informing respondents that a system was differentially private did *not* raise potential user's willingness to share information, no matter which description of DP was presented to the respondent (RQ4).

Taken together, our findings suggest that while (1) respondents do care about the information disclosures against which DP can protect; (2) the likelihood of those disclosures influences respondent's willingness to share; and (3) different in-the-wild descriptions of DP influence respondents perception of the likelihood of those disclosures. However, (4) simply being shown a randomly-selected, in-the-wild description of DP does not increase willingness to share. These results, at first glance, appear to be in tension.

On deeper analysis however, these findings suggest the presence of a misalignment between the information disclosures about which users care and the information disclosures that descriptions of DP address. The probability that a given respondent was (a) shown a description that related to the information disclosures about which they care, and (b) that description influenced enough of their perceptions is likely low.

[1] We also maintain a control group of participants who are not told that their information is protected.

Synthesizing these findings, we posit a novel framework for understanding how end users reason about sharing their data under DP protections. Our framework — and the findings that informed it — offer concrete directions for reformulating DP descriptions to accurately and effectively set user's privacy expectations and increase their comfort when using differentially private systems. Users must either be trained to carefully understand descriptions (as done in [69]) or descriptions should be reformulated to directly communicate how they address the information flows that concern users (e.g., via privacy nutrition-labels [41]). If DP descriptions can be effectively reformulated, our results suggest that users may be significantly more comfortable sharing their information when given differentially private protections.

2 BACKGROUND AND RELATED WORK

In this section, we provide a background on DP and review prior work on communicating privacy to end users, with a specific focus on prior research studying DP-related communications.

Differential Privacy. In the last decade, a growing literature on differentially private algorithms has emerged to address concerns surrounding user-level data privacy. First defined by Dwork et al. [19], DP is a parameterized notion of database privacy that gives a mathematically rigorous worst-case bound on the maximum amount of information that can be learned about any one individual's data through the analysis of a dataset. Formally, a database $D \in \mathcal{D}^n$ is modeled as containing data from n individuals, and DP constrains the change in an algorithm's output caused by changing a single person's data in the database.

Definition 2.1 (Differential Privacy [19]). An algorithm $\mathcal{A} : \mathcal{D}^n \rightarrow \mathcal{R}$ is (ϵ, δ)*-differentially private* if for every pair of databases $D, D' \in \mathcal{D}^n$ that differ in at most one entry, and for every subset of possible outputs $\mathcal{S} \subseteq \mathcal{R}$,

$$\Pr[\mathcal{A}(D) \in \mathcal{S}] \leq \exp(\epsilon) \Pr[\mathcal{A}(D') \in \mathcal{S}] + \delta.$$

DP can be implemented either in the *central model* — where users provide their raw data to a trusted curator for private analysis — or in the *local model*, where users add noise locally to their own data before sharing it for analysis. The central model corresponds to the original DP definition of [19] as presented in Definition 2.1, where an analyst first collects a dataset from users, and then uses specialized DP tools to ensure that the technical requirements of Definition 2.1 are satisfied. The original intended use case for central DP is to enable trusted data analysts who already held sensitive datasets to publish aggregate statistics or reports on their data without violating the privacy of the individuals represented in the data. Central DP is used by e.g., the U.S. Census Bureau [1, 31] and Uber [36, 50] since both require exact user data – the Census Bureau through a constitutional mandate; Uber because data like rider location are necessary for their ride-sharing services.

The local model provides privacy guarantees in the presence of an untrusted curator. Users add noise to their own information (i.e., on their own device) through algorithms that satisfy Definition 2.1 for $n = 1$, and share the privatized output with the curator. Thus, the curator receives only a perturbed and private version of each user's data and never has access to raw user data. Any analysis performed on the noisy data will retain the same DP guarantee due to post-processing [19], so the curator need not use any specialized analysis tools to ensure privacy. Analysts can still make aggregate inferences based on population-level statistics, but will only see noisy information about any individual. Local DP is used by, e.g., Apple [3, 67], Google [24], and Microsoft [16] in settings where user data is stored on-device and the company only requires aggregate information to perform its services.

The possible risks of information disclosure differ substantially between these two models. Since the central model stores user data in a centralized location, data analysts have access to exact user data, along with any other parties who obtain access through legal or illegal means. In the local model, the dataset itself is privatized, so there is no risk of information disclosure through the curator's dataset. In this work, we seek to understand user's perceptions of these possible risks of information disclosure and interrogate the accuracy of those perceptions under both the local and central models of differential privacy.

Privacy Communications. A large body of work has examined how best to explain privacy to end users [22, 35, 44, 63, 66]. This has included creating privacy nutrition labels [41] that clearly delineate to users who may use their information, how their information may be used, and how likely these uses are to occur; designing privacy icons that clearly communicate when and what information is being collected [13, 23, 48]; and developing machine-learning systems that help users negotiate privacy boundaries [61]. Particularly relevant to the work presented here, prior work has also identified best practices for privacy communications: descriptions should be relevant (e.g., include the necessary context for users to make decisions), actionable (e.g., allow the user to make choices), and understandable (e.g., usable, not overloading the user with technical information) [62]. As we discuss in Section 6, our findings suggest that existing DP descriptions fail to satisfy these criteria.

Despite this large body of prior work on privacy communications and the increasing importance of DP, only two pieces of prior work have focused on communicating with users about DP.

Bullek et. al. [9] study how users understand privacy parameters in randomized response, a specific local DP technique. They describe randomized response to users using a virtual, colored spinner; the user would spin the spinner, the outcome of which would indicate if the user should answer the sensitive question truthfully or with the response indicated on the spinner. Our work focuses more broadly on how the information disclosures against which DP protects can influence users' willingness to share, and on how descriptions of DP influence expectations for those disclosures.

Most closely related to our own work, Xiong et al. [69] study how informing users that their information is protected with DP influences their willingness to share different types of information. They study this question in the context of an app that collects medically relevant information, both *low sensitivity* (e.g., gender, height, weight) and *high sensitivity* (e.g., substance use, income level, current medication). They found that promising users DP makes them more willing to share their information (particularly high sensitivity information, which is comparable to the information we consider in this work). However, they found that users struggled to understand descriptions of DP but were more able to understand descriptions that mentioned the implications of information sharing.

Our study builds upon this prior work to more deeply explore (RQ1) which information sharing implications are most concerning to users, and thus should be emphasized when describing DP, (RQ2) how these implications themselves influence users' willingness to share information, and (RQ3) how existing in-the-wild descriptions of DP set their expectations about these information sharing implications. Prior work on user expectations for information sharing [7] notes that there is a lack of work considering how users reason about information sharing under DP. Our work fills this gap. Additionally, we seek to replicate their results through (RQ4), in which we examine how different descriptions of DP themselves influence users willingness to share information.

In addition to primarily focusing on different research questions, our work methodologically differs from the work of Xiong et al. [69] in two ways. First, we derive the descriptions of DP we use as stimuli from a systematic review of 76 in-the-wild descriptions of DP (see Section 5). In contrast, in two of their three experiments Xiong et al. [69] use DP descriptions that were created to explain the definition and/or different aspects (*e.g.*, data perturbation) of DP to users[2]; more similar to our approach, in their third experiment they use the descriptions from four companies that use DP, in addition to their created descriptions. Second, and relatedly, Xiong et al. [69] test whether respondents correctly understand the implications of DP based on the description they were shown before respondents are shown additional survey questions about their willingness to share information. Respondents who did not correctly answer the understanding question(s) were presented with the DP description a second time; if they again did not understand the description, they were excluded from the study. Their results thus have important implications about how users can be educated about DP. On the other hand, this methodology does not offer insight into how users' privacy expectations or willingness to share information might be influenced by encountering descriptions of DP in-the-wild, rather than in a laboratory setting. As further discussed in Section 6, the results of our replication study (RQ4) significantly differ from those of Xiong et al. [69] likely due to the methodological differences in our approaches.

3 SUMMARY OF METHODS

To answer our research questions, we ran two surveys ($n = 2,424$ total), one to address RQ1 and RQ2 and the other to address RQ3 and RQ4. In order to improve the external validity of our work, we use vignette (scenario) surveys [26]. In both our surveys, we use the same two scenarios, which focus on two different potential use cases for DP: protecting people's salaries and protecting people's medical histories. To contextualize the first scenario, we ask respondents to imagine that they work in the banking industry and are approached by a friend on behalf of a salary transparency initiative. In the second scenario, we ask the respondent to imagine that during their next doctor's visit, their doctor asks them if they want to share their medical records with a medical research non-profit, in the name of improving care. The exact wording of these scenarios is shown in Table 1. This table also contains concrete privacy expectations about which we asked respondents in both surveys; we discuss these expectations further in the following sections. We ran both surveys using Amazon Mechanical Turk (MTurk). MTurk has been shown to be representative of American privacy preferences for Americans aged 18-50 who have at least some college education [59].

We present detailed overviews of each survey in Section 4 and Section 5 respectively. Similarly, we present our findings from each survey in Section 4.2 and Section 5.2 respectively. Our full survey instruments can be found in Appendix A. We also present demographic information for our survey respondents in Table 6, which is also located in the appendix. Our procedures were approved by our institutions' ethics review board.

3.1 Limitations

As in all user studies, our study is subject to multiple possible biases. The first is sampling bias. We sample our respondents using MTurk. While prior work shows that MTurk is reasonably representative of the privacy attitudes and experiences of Americans aged 18-50 who have some college education [59], our sample does not capture the experiences of all Americans, especially those older and less educated. Our results should be interpreted in this context. Second, we may have introduced reporting biases through our question design. While we aimed to follow best practices — using cognitive interviews to validate our questionnaire, and offering "other" and "I don't know" response options [57] — respondents may still have mis-reported or failed to report their true perceptions or preferences. Third, while we took steps to improve external validity — sourcing experimental stimuli by rigorously collecting and coalescing in-the-wild DP descriptions and using a vignette survey — our study may have failed to appropriately reflect real-world conditions.

4 IMPACT OF INFORMATION DISCLOSURES (RQ1 & RQ2)

In our first survey, we aimed to answer RQ1 and RQ2. Namely, we wanted to determine if (a) users cared about the kinds of information disclosures against which DP can protect, and (b) users would be more willing to share their sensitive information when the risk of such information disclosures decreased.

Information Disclosures. Contextual integrity (CI) theory – a commonly used framework to explain end-user privacy reasoning – posits that users' privacy decisions depend heavily on information flows – what information is being transmitted to which entities under what privacy expectations [51]. Our scenarios define a set of expected information flows (e.g., salary information moving from the user, to the salary transparency initiative, under some privacy protection – as described further in the following sections). Based on our descriptions, users may have different expectations for whether unexpected information flows (e.g., information being shared with an entity they did not intend to share it with) may occur. We term these unexpected information flows "information disclosures" throughout the remainder of the paper.

As survey one seeks to investigate the role of different types of information disclosures in users' sharing behaviors and survey two seeks to investigate how existing methods of describing DP

[2]The most closely related definition to our work is the DP with implications description: "To respect your personal information privacy and ensure best user experience, the data shared with the app will be processed via the differential privacy (DP) technique. That is, the app company will store your data but only use the aggregated statistics with modification so that your personal information cannot be learned. However, your personal information may be leaked if the company's database is compromised."

Scenario Name	Scenario Description
Salary Scenario	"Imagine that you work in the banking industry. You are friends with a group of other people who work in banking companies in your city. One of your friends is part of a transparency initiative that is trying to publish general statistics about pay in the banking industry. As part of this initiative, they have asked everyone in the group to share their salaries and job titles using an online web form on the initiative's website."
Medical Scenario	"Imagine that during your next doctor's visit, your primary care doctor informs you that they are part of a non-profit organization trying to push the boundaries of medical research. This non-profit is asking patients around the country to share their medical records, which will be used to help medical research on improving treatment options and patient care. Your doctor, with your permission, can facilitate the non-profit getting the information they need."

Expectation Name	Expectation Description	Ground Truth: Local	Ground Truth: Central
Hack	"A criminal or foreign government that hacks the transparency initiative could learn my salary and job title"	False	True
	"A criminal or foreign government that hacks the non-profit could learn my medical history"		
Law Enforcement	"A law enforcement organization could access my salary and job title with a court order requesting this data from the initiative"	False	True
	"A law enforcement organization could access my medical history with a court order requesting this data from the non-profit"		
Organization	"My friend will not be able to learn my salary and job title"	True	False
	"The contents of my medical record will be stored only by my doctor's office, and will not be stored by the non-profit"		
Data Analyst	"A data analyst working on the salary transparency initiative could learn my exact salary and job title"	False	True
	"A data analyst working for the non-profit would be able to see my exact medical history"		
Graphs	"Graphs or informational charts created using information given to the salary transparency initiative could reveal my salary and job title."	False	False
	"Graphs or informational charts created using information given to the non-profit could reveal my medical history."		
Share	"Data that the salary transparency initiative shares with other organizations doing salary research could reveal my salary and job title"	False	True
	"Data that the non-profit shares with other organizations doing medical research could reveal my medical history"		

Table 1: Scenarios (top) and information disclosure expectations (bottom) used in both survey one and survey two.

influence user expectations for information disclosure, both surveys address the same potential information disclosures that could occur in either scenario.

While we would like users to tell us about the kinds of disclosure that concerns them, prior work has shown that users often do not have good mental models for privacy tools [2, 42, 43]. To compensate for this, we leveraged prior work [10, 32, 37, 47, 52, 68] on both DP and on user privacy concerns to create a preliminary list of information disclosures about which a user might care. Our list included the following kinds of information disclosure (names in italics): *(Hack)* Could a criminal organization or foreign government access the respondent's information by hacking the organization holding the information?; *(Law Enforcement)* Could a law enforcement organization access the respondent's information with a court-issued warrant?; *(Organization)* Could the organization collecting the information (or their representative) access the respondent's information?; *(Data Analyst)* Could a data analyst working within the organization access the respondent's information?; *(Graphs)* Could graphs or informational charts created by the organization be used to learn the respondent's information?; and *(Share)* Could the collected information be shared with another organization such that the other organization could access the respondent's information?

While some of these questions are redundant from a technical perspective (e.g., *hack* and *law enforcement*), we chose these questions to be representative of real data-privacy concerns that potential users might have, since prior work finds that a key part of users' reasoning about privacy is how appropriate they consider different information flows [51].

4.1 Methodology: Survey One

To ensure that we had not missed information disclosures about which users were concerned, we first conducted five cognitive

interviews[3] and offered survey respondents the opportunity to list other information disclosures about which they cared. Fewer than 2% entered ag disclosure not captured in our list. As such, we use the above list of information disclosures throughout this work. We present the descriptions of these expectations in Table 1.

In Table 1, we also indicate the ground truth for each of these information disclosures in both the local and central model of DP. Both central and local DP protect against information disclosure through graphs, as this is the core privacy guarantee of DP. The central model aggregates raw user data into a centralized database that can potentially be accessed by the data analyst, employees of the organization, entities that hack the organization, law enforcement (with proper court orders), and partner organizations with whom the dataset is shared. In the local model, the aggregated dataset contains only DP versions of user data, so information disclosure would not occur even if the dataset is accessed by these entities.

We stress that we are considering a "typical" DP deployment and acknowledge that there are deployments for which our ground truths are not correct. For instance, we indicate that a data analyst would be able to learn a potential user's exact information in the central model. However, Uber deployed central DP specifically to protect users' information against curious data analysts. As it is impossible to account for all possible system parameters and design options, we derive our ground truth from the most simple setup.

Questionnaire. Each respondent was randomly assigned to either the salary scenario or the medical scenario described above. Then, each respondent was asked to indicate which of the information disclosures, described above, they would want to better understand before sharing their information. Additionally, they were given the option of adding any additional disclosures about which they would want additional information. For each information disclosure event that the respondent indicated they would want to better understand, the respondent was presented with one of the following explicit risks, chosen at random:

(1) there is no risk of this information disclosure,
(2) the risk of this information disclosure is the same as the chance that your bank account will be compromised (accessed by a person who you did not intend to gain access to) as part of a data breach in the next year, and
(3) the risk of this information disclosure is higher than the chance that your bank account will be compromised (accessed by a person who you did not intend to gain access to) as part of a data breach in the next year.

We set expectations in this way because (a) prior work on how humans interpret numbers and risk suggests that reference events of a similar type improve risk comprehension [25, 40, 60, 64], and (b) prior work shows that users have concrete estimates for the likelihood of bank account compromise, a frequently discussed security event [5, 38, 65]. Each respondent was then asked if they would be willing to share their information with the initiative. Additionally, they were asked to describe why they would or would not be willing to share their data.

[3]Cognitive interviewing is a survey methodology technique in which participants think aloud as they answer a survey [57]. Cognitive interviews are used to verify that potential respondents understand the survey questions and no answer choices are missing. We conducted interviews until no new survey protocol corrections emerged.

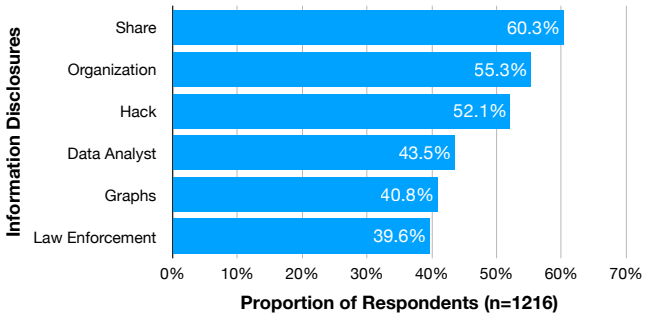


Figure 1: Proportion of respondents who care about each potential information disclosure.

Finally, each survey concluded with a battery of demographic questions, including a measurement of internet skill using an existing validated measure [28], as prior work suggests that internet skill is among the most relevant constructs to control for in privacy studies [29, 30, 56, 58]. The complete survey is in Appendix A.1.

Sample. We surveyed 1,216 U.S. Amazon Mechanical Turk workers. These workers were split evenly between the two survey scenarios. To ensure high quality responses, we required that respondents have at least a 95% approval rating [55]. The demographics of our sample are reported in Table A in the appendix.

Analysis. To answer RQ1, we conduct a descriptive analysis, reporting the proportion of respondents who were concerned about each potential information disclosure event; when reporting differences between proportions of respondents who report concern, we use χ^2 proportion tests to validate that the differences between proportions are significant. To answer RQ2, we build six logistic regression models, one for each potential information disclosure event. In each model, the dependent variable (DV) is whether the respondent is willing to share their information, the independent variable (IV) of interest is the level of risk that the respondent was told of the information disclosure occurring, and the control IVs are the scenario type and the respondents' internet skill score. We report the odds ratios (the exponentiated regression coefficients) with 95% confidence intervals, and p-values for each IV in the model.

In Section 4.2, we contextualize a subset of our results using open-text responses participants provided to describe their sharing decisions. These responses were analyzed through open-coding by a member of the research team with qualitative research experience. As these responses are not offered as primary research artifacts, we do not double code this data nor provide intercoder agreement statistics, per best-practice guidelines outlined in [45].

4.2 Information Disclosure Results

Here, we detail the results of our analyses of survey one.

RQ1: What Information Disclosures Concern Users? The goal of DP is to protect user information against disclosure to various entities. Thus, we investigate whether users care about potential information disclosures to different entities against which DP can protect (see Table 1 for information disclosures and Section 4.1 for the source of these disclosures).

Variable	*Hack*		*Law Enforcement*		*Organization*		*Data Analyst*		*Graphs*		*Share*	
	OR/CI	p-value	OR/CI	p-value	OR/CI	p-value	OR/CI	p-value	OR/CI	p-value	OR/CI	p-value
Low Risk	**1.91** [1.27, 2.88]	**< 0.01****	0.87 [0.54, 1.39]	0.55	**1.63** [1.12, 2.38]	**0.01***	0.92 [0.59, 1.44]	0.72	1.42 [0.89, 2.27]	0.14	1.25 [0.85, 1.84]	0.27
No Risk	**2.97** [1.98, 4.49]	**< 0.01*****	**2.07** [1.32, 3.27]	**< 0.01****	**1.61** [1.1, 2.35]	**0.01***	1.23 [0.79, 1.9]	0.36	1.48 [0.93, 2.37]	0.10	**1.97** [1.35, 2.88]	**< 0.01*****
Salary Scenario	**1.44** [1.04, 2]	**0.03***	1.35 [0.92, 1.98]	0.12	1.20 [0.89, 1.64]	0.23	1.37 [0.96, 1.96]	0.09	1.12 [0.77, 1.64]	0.54	1.04 [0.76, 1.42]	0.81
Internet Score	1.04 [0.86, 1.25]	0.68	1.01 [0.82, 1.24]	0.92	0.98 [0.82, 1.16]	0.79	**1.24** [1.01, 1.52]	**0.04***	1.05 [0.86, 1.28]	0.66	1.01 [0.86, 1.2]	0.87

Table 2: Effect of expectations about the probability of a disclosure on respondent willingness to share. This Table is constructed using data from Survey One. Each logistic regression model constructed only for respondents who cared about that type of disclosure, comparing against the High Risk condition. Table shows odds ratio (OR), 95% confidence intervals for the odds ratios (shown in brackets), and p-values, where * indicates $p < 0.05$, ** indicates $p < 0.01$, and * indicates $p < 0.001$.**

We find that the most respondents — 60.3% — care about information disclosures to third-parties (Share). Over half (55.3%) care about disclosures to the person or organization running the initiative to which they contributed their information, while 52.1% care about disclosing their information to an entity that hacks the organization to which they contributed their information. Fewer, 43.5%, care about whether a data analyst working at the organization might be able to learn their private information or whether graphs created using their information might disclose their private information (40.8%). Finally, 39.8% of respondents care whether law enforcement might be able to access their information using a court order. We visualize these findings in Figure 1.

It is interesting to note that nearly 20% more respondents ($\chi^2 = 34.54$, $p < 0.001$) cared about their information being disclosed to a third party vs. being disclosed through graphs created using their information. Similarly, 11.8% more respondents ($\chi^2 = 22.32$, $p < 0.001$) cared about their information being disclosed to the organization running the initiative vs. being disclosed specifically to a data analyst at the organization. Open-answer responses offer some insight into this difference. When asked why they would (or would not) be willing to share their information, after they were told the risk of the disclosures they indicated they cared about, many respondents indicated that they believed in the cause of the initiative and wanted to contribute to their analysis/research. For example, one respondent said, "I trust the non-profit organization to handle my information responsibly and to use it for the positive research purpose that they claim they will be using it for." Thus, respondents may care less about disclosures that occur through "appropriate information flows" [51], in which user's information is being used in the way they expect: e.g., to benefit salary transparency or medical research through data analysis and the generation of graphs.

RQ2: How Does the Probability of Information Disclosure Influence Sharing? Next, we investigate whether setting respondents' expectations about the information disclosures that concerned them influences their reported willingness to share information. The results are shown in Table 2.

Among respondents who care about their information being hacked by a criminal organization or foreign government, those who were told that the risk of this disclosure occurring is about the same as the risk of having their bank account compromised ("Low Risk" in Table 2) are nearly twice (O.R. = 1.91, $p < 0.01$) as likely to share their information as compared to respondents who were told that this risk is greater than the risk of having their bank account compromised. Those who are told there is no risk of their information being disclosed through a hack were nearly three times as likely to share (O.R. = 2.97, $p < 0.01$). Respondents who care about their information being hacked were also more willing to share salary information than medical information.

Respondents that care about information disclosure to the organization running the initiative to which they might contribute their information were > 60% more willing to share if the risk of their information being disclosed to the organization was lower (Low Risk: O.R. = 1.63, $p = 0.01$; No Risk: O.R. = 1.61, $p = 0.01$).

On the other hand, respondents who care about whether law enforcement would be able to access their information with a court order and respondents who care about whether their information might be disclosed to a third-party are both more likely to share their information only if they are told there is no risk of their information being disclosed to these entities. Respondents that are told there is no such risk are about twice as likely to share their information (Law Enforcement: O.R. = 2.07, $p < 0.01$; Share: O.R. = 1.97, $p < 0.01$). Being told there is a low risk instead of a high risk of disclosure has no significant effect on their willingness to share.

We hypothesize that respondents show a graduated response to the risk of information disclosures to the organization running the initiative to which they might contribute their information because it is appropriate for this organization to have their information. Similarly, we hypothesize that respondents show a graduated response to the risk of hacks because information disclosures resulting from hacks are unintentional on the part of the organization. On the other hand, the organization purposefully choosing to share information they contributed to the initiative with a third party or with law enforcement, even with a court order, may feel incongruent with the purpose for which they shared their information.

Finally, we find that the probability of disclosure to data analysts or through graphs has no effect on willingness to share, even among respondents who care about those information. We hypothesize that those respondents who are motivated by the altruistic

Theme	Description
Unsubstantial	"Differential privacy is the gold standard in data privacy and protection and is widely recognized as the strongest guarantee of privacy available."
Techniques	"Differential Privacy injects statistical noise into collected data in a way that protects privacy without significantly changing conclusions."
Enables	"Differential Privacy allows analysts to learn useful information from large amounts of data without compromising an individual's privacy."
Trust	"Differential privacy is a novel, mathematical technique to preserve privacy which is used by companies like Apple and Uber."
Risk	"Differential privacy protects a user's identity and the specifics of their data, meaning individuals incur almost no risk by joining the dataset."
Technical	"Differential privacy ensures that the removal or addition of a single database item does not (substantially) affect the outcome of any analysis. It follows that no risk is incurred by joining the database, providing a mathematically rigorous means of coping with the fact that distributional information may be disclosive." [17]

Table 3: Descriptions of DP synthesized from the six main themes present in our collection of 76 in-the-wild DP descriptions.

goals specified in the scenarios may be willing to share their information regardless of the risk of disclosure occurring through these information flows, which are arguably the most appropriate information flows we examine, while those who are not compelled by the goals of the organizations described in the scenarios are similarly unwilling to share their information regardless of this risk. For example, one respondent who cared about information disclosure to both of these entities, and was told there was no risk of disclosure to a data analyst and low risk of disclosure through a graph said, "Unfortunately, I do not see enough of a benefit for me to take the risk of sharing my personal information. I absolutely do not want such personal info being leaked out." On the other hand, a respondent who cared about disclosure to a data analyst and was told the risk of this disclosure was higher than the risk that their bank account would be compromised commented that they would be willing to share their information, "because it's for good research, and I'm getting too old to worry about who sees my medical record. I anticipate I will have *many* doctors, nurses, lab techs, etc involved in my medical record before too long."

We note that respondents who cared about information disclosures to data analysts with higher internet scores were more likely to report being willing to share their information. As technologically savvy respondents, they may have had a clearer mental model of the data analysis process and therefore understood that data analysts typically have complete access to user information. As such, they may be more forgiving toward any approach that aims to reduce this level of access, even given the relatively high risk of an information disclosure.

5 EXPECTATIONS & WILLINGNESS TO SHARE UNDER DP (RQ3 & RQ4)

Next – via a second survey – we explore how DP influences privacy expectations (RQ3) and intent to share information (RQ4).

Descriptions of Differential Privacy. In order to answer these research questions, we needed to describe DP to respondents in our surveys. However, there is no standard description of DP we can use. Because we want to ask our research questions in a realistic context, we seek to describe DP to our respondents in the same way they might encounter DP in-the-wild.

To determine how DP is described in-the-wild, we conducted a systematic search for publicly available descriptions of DP using keywords such as "differential privacy," "formal privacy," "privacy guarantee," and "census privacy." We used both Google search and searched within the past five years (2014-2019) of content in large media venues. We continued searching until new search results stopped appearing. We put special focus on collecting descriptions used by industry and in the media coverage, as these descriptions are the ones that an uninformed consumer would be most likely to encounter. We performed this search and data collection in December 2019.[4] In total, we collected 76 descriptions of DP: 36 from industry, 30 from media outlets, and 10 from the academic literature.

The industry descriptions primarily came from companies that use DP, including Google, Apple, Microsoft, and Uber, as well as smaller start-ups and an investment firm. We also gathered multiple descriptions from the U.S. Census Bureau regarding the use of DP in the 2020 Census. The media descriptions were from large, mainstream media outlets, such as The New York Times, Fox News, The Washington Post, The Guardian, and Tech-Crunch. The academic descriptions were collected from some of the most-cited papers and books on DP, e.g., [17]. As DP is an active area of research, these ten academic descriptions are clearly not comprehensive, but serve as a representative example of academic descriptions.

The research team employed affinity diagramming [8] to extract the main themes of these widely varying descriptions of differential privacy. In affinity diagramming, the research team collaboratively sorts pieces of content — in our case the descriptions of DP — into themes based on affinity, with each researcher iterating over the affinity diagram at least twice until consensus was reached on appropriate categorization. This analysis resulted in the identification of six main themes (names in bold): **(Unsubstantial)** claims that DP is the best notion of privacy; **(Techniques)** explanations

[4]Since we conducted this survey, more companies and organizations have started adopting and publicly writing about DP. As such, our dataset is no longer comprehensive. Because this data collection informed the design of our survey, we choose not to incorporate the newer descriptions into our dataset.

Variable	*Hack*		*Law Enforcement*		*Organization*		*Data Analyst*		*Graphs*		*Share*	
	OR/CI	p-value	OR/CI	p-value	OR/CI	p-value	OR/CI	p-value	OR/CI	p-value	OR/CI	p-value
Description: *Unsubstantial*	**1.94** [1.16, 3.29]	**0.01***	1.10 [0.65, 1.86]	0.72	1.13 [0.73, 1.75]	0.59	1.71 [0.92, 3.27]	0.10	**1.64** [1.01, 2.67]	**0.05***	1.68 [0.99, 2.88]	0.06
Description: *Techniques*	**1.96** [1.17, 3.33]	**0.01***	1.21 [0.72, 2.03]	0.47	1.43 [0.93, 2.22]	0.10	**2.40** [1.33, 4.5]	**< 0.01****	**2.15** [1.34, 3.5]	**< 0.01****	**2.22** [1.33, 3.77]	**< 0.01****
Description: *Enables*	1.60 [0.95, 2.73]	0.08	1.05 [0.63, 1.77]	0.84	1.40 [0.91, 2.16]	0.13	**2.06** [1.13, 3.88]	**0.02***	**1.76** [1.09, 2.87]	**0.02***	1.69 [1, 2.9]	0.05
Description: *Trust*	**1.86** [1.11, 3.17]	**0.02***	1.04 [0.61, 1.76]	0.89	1.43 [0.92, 2.22]	0.11	**1.99** [1.08, 3.78]	**0.03***	1.38 [0.84, 2.28]	0.20	1.19 [0.68, 2.09]	0.55
Description: *Risk*	**2.58** [1.57, 4.33]	**< 0.01*****	**1.86** [1.15, 3.05]	**0.01***	1.43 [0.93, 2.20]	0.10	**2.46** [1.37, 4.59]	**< 0.01****	**2.40** [1.50, 3.88]	**< 0.01*****	**2.27** [1.37, 3.84]	**< 0.01****
Description: *Technical*	1.56 [0.92, 2.69]	0.10	1.02 [0.60, 1.73]	0.95	1.38 [0.89, 2.14]	0.15	**2.30** [1.26, 4.33]	**0.01****	**1.70** [1.04, 2.79]	**0.03***	**1.90** [1.12, 3.25]	**0.02***
Salary Scenario	**1.32** [1.02, 1.71]	**0.04***	0.80 [0.61, 1.05]	0.11	**1.29** [1.03, 1.63]	**0.03***	0.75 [0.56, 1.01]	0.06	1.23 [0.96, 1.57]	0.10	1.14 [0.88, 1.49]	0.31
Internet Score	**1.17** [1.01, 1.36]	**0.04***	**1.25** [1.06, 1.46]	**0.01****	1.02 [0.89, 1.17]	0.78	1.05 [0.89, 1.25]	0.54	1.07 [0.93, 1.23]	0.36	1.01 [0.87, 1.18]	0.85

Table 4: Effect of DP descriptions on respondent's perception of the likelihood that their information will be disclosed through a particular information flow. All models are logistic regressions constructed using data from Survey Two. See Table 2 for detailed legend.

that briefly summarize the methods used to create differentially private summary statistics, usually focusing on statistical noise; **(Enables)** statements that attempt to capture the types of applications that DP makes possible; **(Trust)** descriptions that focus on the well known organizations and companies that have recently started using DP; **(Risk)** statements that highlight the data-privacy risks that an individual incurs when allowing their information to be part of a differentially private system; and **(Technical)** highly technical explanations using dense, mathematical language.

Many of the descriptions we gathered touch on more than one of these main themes. For instance, documents prepared by the U.S. Census Bureau state, "Differential privacy allows us to inject a precisely calibrated amount of noise into the data to control the privacy risk of any calculation or statistic" [31]. This description touches on the *techniques* theme and the *risk* theme, while also using *technical* language like "calculation or statistic" that may be unnatural to non-experts. The New York Times provides a description that is another combination of the main themes, writing, "[o]ne example, differential privacy, is already used by Apple, Google and even the U.S. Census Bureau to limit the amount of personal information that is shared with an organization while still allowing it to make useful inferences from the data" [12] This description contains elements of both the *trust* theme and the *enables* theme.

We note that most descriptions we gathered did nothing to distinguish between the central model and the local model. Indeed, we found that determining if an industry system was in the local or central model generally required looking at technical documentation. The descriptions provided by media coverage also generally did not include any indication as to the model of system being described.

5.1 Methodology: Survey Two

After collecting and analyzing the descriptions of DP used in practice, we distilled six descriptions of DP that were representative of the descriptions in each of these themes. We present these descriptions in Table 3. Each of these descriptions is a synthetic creation meant to be representative of the real descriptions we found, with the exception of the *technical* description, which was taken from [17]. We chose to create new descriptions (rather than selecting a representative example) in order to free the description from the surrounding context, make them consistent in their structure, and have each description focus on only one theme. Just like the descriptions we observed in-the-wild, our descriptions included no indication as to whether the system they described was in the local or central model.

Questionnaire. First, as in survey one, each respondent was randomly presented one of the two scenarios. Each respondent was then randomized to either a control condition, where there was no mention of privacy protection, or shown one of the DP descriptions. In the later conditions, the scenario was followed by the description, "To reduce the intrusion into personal privacy, the [organization] will use a technique called differential privacy. [differential privacy description]," where the description presented was sampled with equal weight from Table 3. Respondents were then asked the following questions: First, they were asked if they would be willing to share their data. Next, each respondent was asked to share their concrete privacy expectations by reporting whether they believed the expectations described in Table 1 (e.g., "A criminal or foreign government that hacks the transparency initiative could learn my salary and job title.") were true or false. All questions included an "I don't know" option. Finally, we included the same demographics questions as above. The complete survey is in Appendix A.2.

Sample. We surveyed 1,208 Amazon Mechanical Turk workers following the same screening requirements as in survey one, as described in Section 4. Workers were split evenly between the two survey scenarios and equally between the seven description conditions (six descriptions of DP and the baseline).

Analysis. To answer RQ3, we construct six logistic regression models [5], one for each information disclosure event. The DV is whether the respondent reported that they thought the given information disclosure would occur, the IV of interest is the description they were shown (a categorical variable with the control – no description shown – as the baseline), the control IVs are, as in Section 4, whether the scenario was the salary scenario and the respondent's internet skill. To answer RQ4, we construct a single logistic regression model. The DV is whether the respondent reported that they would be willing to share their information, the IV of interest is the description they were shown (coded as above), and the control IVs are the same as above.

5.2 Results: Responses to Descriptions of DP

Here, we detail the results of our analyses of survey two.

RQ3: How Do Differential Privacy Descriptions Affect Privacy Expectations? We divide our results for RQ3 into two parts. First, we detail our findings regarding the way descriptions of DP *increase* respondents' privacy expectations (see Table 4). Second, we investigate if the descriptions *correctly* set respondents' privacy expectations, with respect to the ground truth privacy properties of typical local and central DP deployments (see Figure 2).

Descriptions Increasing Expectations. Overall, we find that each description of DP that we tested increased respondents' privacy expectations for at least one of the disclosure risks. However, different descriptions increased different privacy expectations.

First, we found that none of the descriptions significantly changed respondents' expectations when it came to disclosing their information to the organization soliciting their information it's representative. Respondents had higher privacy expectations in the salary scenario than in the medical scenario, indicating that the slightly different wording of these expectations may have had an effect on respondent expectations. We verified that this wording did not interfere with our main finding (that no descriptions increased user's expectations of the *Organization* disclosure) by re-building our models on each dataset separately; we found the same results.

Four of the descriptions do, however, influence respondents' perceptions regarding whether their information could be disclosed through a hack. Respondents who were shown the *Unsubstantial*, *Techniques*, and *Trust* explanations were nearly two times more likely (*Unsubstantial:* O.R. = 1.94, $p = 0.01$, *Techniques:* O.R. = 1.96, $p = 0.01$, *Trust:* O.R. = 1.86, $p = 0.02$) to think their information would not be disclosed through a hack. Those who were shown the *Risk* description were even more likely (O.R. = 2.58, $p < 0.01$) to think their information could not be disclosed through a hack. Users may see preventing hacks as one of the key roles of security and privacy technologies, as these results indicate that they expect such protection from the gold standard techniques (*Unsubstantial*) and those used by major companies (*Trust*). The *Risk* description directly addresses this potential information disclosure, so it is unsurprising that it raised privacy expectations regarding hacks. Finally, respondents may have gathered that injecting statistical noise (*Techniques*) would protect their information against hacks, as it does in practice.

Only the *Risk* description significantly influenced respondents' perceptions of whether their information would be disclosed to law enforcement as the result of a court order: those who saw the *Risk* description were nearly two times more likely (O.R. 1.86, $p = 0.01$) to think their information would not be disclosed to law enforcement. Interestingly, this indicates that users may see information sharing with law enforcement as a risk, rather than an information flow that is appropriate and necessary to protect society.

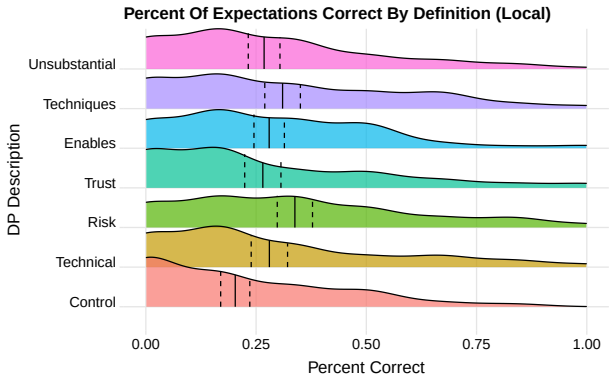
All of the descriptions aside from the *Unsubstantial* description increase the likelihood that respondents think their information would be secure against disclosures to a data analyst, while all but the *Trust* description increase the likelihood that respondents think their information would not be disclosed through graphs or charts made using their information. It may be that the *Unsubstantial* description did not raise respondents' expectations for disclosure to data analysts because users are unfamiliar with the notion that data analysts could accomplish their job without seeing user information — users may expect even "gold standard" techniques to disclose information in this way. Similarly, users may be unfamiliar with the idea that tech companies create graphs and charts, as most of these releases are not customer facing. Therefore, it would not be assumed that such techniques could protect user information.

The *Techniques*, *Risk*, and *Technical* descriptions all increase the likelihood that respondents think their information could not be shared with another organization (*Techniques:* O.R. 2.22, $p < 0.01$, *Risk:* O.R. 2.27, $p < 0.01$, *Technical:* O.R. 1.90, $p = 0.02$). As above, users may have gathered that the injection of statistical noise described in the *Techniques* description would prevent this information disclosure. Additionally, both the *Risks* and *Technical* descriptions speak to the risk of joining the dataset. As indicated in our results for RQ1, a large number of respondents care about their information being disclosed to a third party. This may be a primary "risk" in their mind, which the descriptions suggest that they would be secure against.

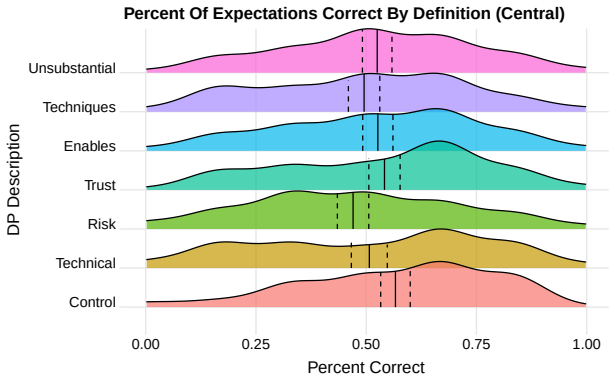
Descriptions Setting Expectations Correctly. Not all DP techniques reduce the likelihood of all potential information disclosures. It is critical that descriptions of DP are used to set users' expectations correctly, not only raise expectations. This is especially important in DP; a potential user encountering a description of a DP may set their expectations as though the system offers local DP, only to discover later that their information was more vulnerable because the deployment used central DP (see Table 1 for the ground truth we consider under both local and central DP).

Revisiting our findings above, we note that local DP provides protection against all of the information disclosure risks about which we asked. This is because under local DP the curator never has access to the unperturbed data (and therefore cannot accidentally or intentionally disclose information). As such, increased expectations mean more correct expectations under local DP. As we saw above, each of the descriptions does increase some — but not all — expectations. This means that the descriptions are not only raising user

[5] We note that these models are not corrected for multiple testing in line with Benjamini and Hochberg [6], which suggests such correction only for models with a large number of DVs.



(a) **Distribution of respondent correctness about information disclosures under local DP. The x-axis shows percentage of disclosures for which their expectations were correct under local DP.**



(b) **Distribution of respondent correctness about information disclosures under central DP. The x-axis shows percentage of disclosures for which their expectations were correct under central DP.**

Figure 2: Influence of in-the-wild DP descriptions on respondent expectations for information disclosures by DP model. Dashed lines are 95% confidence intervals of the distribution's mean. "I don't know" not included as correct.

expectations for differentially private systems, but setting those expectations more accurately for local privacy.

In central DP, on the other hand, the curator has access to users' raw information. In this model, the curator is responsible for injecting statistical noise into any aggregations that are released for public consumption. Because the curator has access to raw information, a typical deployment would be able to disclose sensitive information in all of the listed ways, with the exception of *Graphs*. Thus, the descriptions that raise privacy expectations related to *Hack*, *Law Enforcement*, *Data Analyst*, and *Share* disclosures are actually misleading users in the central model.

Finally, we also consider the aggregate effect of the descriptions we study on the accuracy of respondents' privacy expectations (see Figure 2). We find that respondents' expectations of DP are more in line with the central model than the local model (ie. they have lower privacy expectations). Specifically, respondents had correctly set expectations for less than half of the information disclosure risks under local DP, while roughly half of their expectations were set correctly for the central model. More importantly, we note that users' privacy expectations are poorly aligned with *both* local and central DP. This indicates that users have no coherent mental model of the data collection process, as many of the privacy expectations about which we ask are equivalent from a technical perspective.

RQ4: How Do In-The-Wild Descriptions of Differential Privacy Affect Sharing? When analyzing the results of our second survey, we find that respondents who were told that their information would be protected by DP techniques, as described by one of six different descriptions of those techniques, were no more likely to report that they would share their information in either scenario (Table 5). We also note that the descriptions did not decrease the likelihood that respondents would be willing to share their information. Respondents were, however, more likely (O.R. = 1.67, $p < 0.01$) to share their information in the salary scenario than in the medical scenario, in line with prior work suggesting that medical information is particularly sensitive [34].

Variable	Odds Ratio	CI	p-value
Description: *Unsubstantial*	1.22	[0.79, 1.88]	0.37
Description: *Techniques*	0.96	[0.62, 1.47]	0.83
Description: *Enables*	1.48	[0.96, 2.29]	0.08
Description: *Trust*	1.08	[0.7, 1.67]	0.72
Description: *Risk*	1.37	[0.89, 2.12]	0.15
Description: *Technical*	0.94	[0.61, 1.45]	0.77
Salary Scenario	**1.67**	**[1.32, 2.1]**	$<$ **0.01*****
Internet Score	1.09	[0.95, 1.25]	0.2

Table 5: Effect of DP descriptions on respondents' likelihood of being willing to share information. See Table 2 for detailed legend.

We note that this finding contradicts the findings of [69], who found that DP increased respondents' willingness to share high sensitivity information. We note that (a) the ways in which we describe DP and (b) the context in which we elicit responses are different. As discussed in Section 2, the descriptions used in this prior work were significantly longer and not necessarily representative of in-the-wild descriptions from which we derived our descriptions, and the methodology of prior work also involved mechanisms to ensure respondents correctly understood the privacy guarantees detailed in the descriptions.

6 DISCUSSION

Summary of Findings. Our surveys indicate that (RQ1) users care about the kind of information disclosures against which DP can protect, and (RQ2) users' willingness to share information is significantly related to the degree of risk of most information disclosures occurring. However, the risk of disclosures occurring through the two information flows that might seem most appropriate [51] to users given our scenarios related to a salary transparency initiative and a medical research initiative — disclosures through graphs or to a data analyst — did not significantly relate to users' willingness

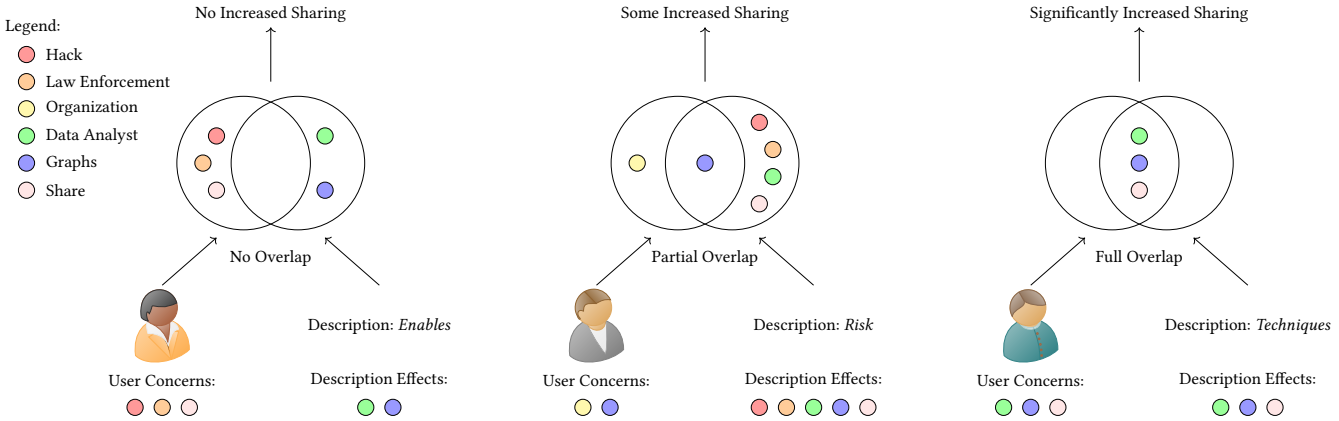


Figure 3: Visualization of our framework for reasoning about the impact of DP descriptions on users' willingness to share information. Colored dots under users represent information flows about which that user cares. Colored dots under descriptions represent the information flows for which the description raises expectations. We imagine a potential user with some prior set of information disclosures about which they are concerned. When asked if they are willing to share their information with a differentially private system, the user is given a brief description of DP. Our results suggest that a user's willingness to share information is not simply a function of how this description raises their expectations, but also a function of their prior concerns. Specifically, the description may raise the user's expectations for information disclosures about which the user was not concerned. Thus, we speculate that the degree to which the user's expectations overlap with the effects of a description will be an important determining factor in a user's willingness to share.

to share. This is noteworthy, as ensuring privacy in graphs and informational charts is a common motivating example of DP, is the only information disclosure protected against by both local and central DP, and at least one current deployment of DP is focused on protecting user information from data analysts [50].

We also find that in-the-wild descriptions of DP have a substantial impact on user privacy expectations (RQ3), but not user willingness to share (RQ4). Descriptions of DP that focus on different themes raise privacy expectations for information flows. However, this can be a double-edged sword, as raising expectations can also mislead users about the privacy properties of a system.

Novel framework for reasoning about the impact of descriptions. Upon first inspection, there appears to be a contradiction embedded in our results: we established that respondents care about information disclosures relevant to DP and are (in some cases) more willing to share their information when they are assured that these information disclosures will not occur. But, we found that offering respondents DP did *not* increase their willingness to share information, no matter the description. At first glance, these results might seem to indicate that respondents did not understand the descriptions at all. However, the results in Table 4 show that respondents had higher privacy expectations when presented with some descriptions. One would expect that these higher expectations would lead to higher willingness to share, in line with our first results.

To resolve this tension, we recall that not every respondent cared about every kind of information disclosure (Section 4.2). While many respondents cared about each kind of disclosure, none of the information disclosures were important to more than 60% of respondents. Thus, there was almost certainly misalignment between the disclosures that mattered to a given respondent and the disclosures that were influenced by the DP description they were shown. For instance, imagine a potential user cared about the *Share* expectation, but was presented with the *Trust* description. This potential user's higher expectations for *Hack* and *Data analyst* disclosures would likely do little to raise their willingness to share.

These results suggest a framework for understanding how descriptions of DP influence a user's willingness to share information (visualized in Figure 3). When users encounter a differentially private system, they already have privacy preferences and concerns. When a user sees a description of DP, the user's expectations about certain information flows may increase. If the ways in which their privacy expectations increase aligns with the types of information disclosure about which they are concerned, they may be more likely to share their private information. A key takeaway from this framework is that a clear and concise description of DP may not be enough to raise user's willingness to share. Instead, it is important that a description speaks to users' concerns directly and be tailored to address those concerns, as we discuss below.

Need for new descriptions. It is very evident from our results that the way DP is described in-the-wild is insufficient to help users make informed decisions. There is no consistency or standardization in the language organizations use. Thus, characterizing the way users might see DP described required us to identify the six descriptive themes used in this work. The themes present in these descriptions seem to haphazardly raise users' expectations. This is especially concerning given the differences between local and central DP; if descriptions are not carefully tailored to the model, they may mislead users about the privacy properties of the system. Indeed, Figure 2 shows that the existing descriptions of DP do little to correctly set expectations, no matter the deployment model.

We note that the simple descriptions that we showed respondents in our surveys are not completely ineffective or without use. For instance, using our *Risk* description may be appropriate for a local

DP deployment as it raised expectations broadly. However, because these descriptions do nothing to increase participation, they may not achieve the goals of system designers.

There are two main alternatives for improving the state of DP descriptions. First, one could take the approach of Xiong et al. [69], carefully constructing descriptions of DP and training users to understand those descriptions. However, Xiong et. al.'s results indicate that such an approach is difficult: a significant number of users were unable to correctly answer test questions about DP after viewing carefully crafted descriptions. An alternative approach, which prior work on privacy beyond DP suggests may be particularly effective [62], is to explicitly inform users about the risks posed to their information. For instance, a description of a central DP system might specify that information will not be leaked through any graphs or informational chats, but could still be leaked to the other entities listed above. This would be similar to the privacy nutrition labels proposed by Kelley et al. [41]. Such descriptions of DP could allow users to make an informed information sharing decision without requiring them to build a comprehensive mental model of the technical details of DP techniques. That said, technical details and parameter choices for DP deployment (e.g., the value of ϵ) have important implications for user privacy and, as such, future work should also explore how best to communicate these technical nuances in meaningful ways.

7 CONCLUSION

In this work we studied DP from the user's perspective, focusing on how users' privacy expectations relate to DP as they are likely to encounter it in-the-wild. We showed that the privacy concerns about which users care can be addressed by DP, but the varied ways in which DP is described set user expectations in a haphazard, and often inaccurate, manner. Our results indicate that the interaction between user's intrinsic privacy concerns and the ways in which descriptions of DP set user expectations informs a user's willingness to share their information under differentially private guarantees. Our work posits a novel framework for understanding this interplay and suggests concrete directions for developing better descriptions of DP that directly and accurately address user privacy concerns.

ACKNOWLEDGMENTS

The first author was supported in part by The Defense Advanced Research Projects Agency (grant number W911NF-21-1-0371), NSF grants CNS-1850187 and CNS-1942772 (CAREER), a Mozilla Research Grant, and a JPMorgan Chase Faculty Research Award. Part of this work was completed while the first author was at Georgia Institute of Technology. The second author is supported by the National Science Foundation under Grant #2030859 to the Computing Research Association for the CIFellows Project and The Defense Advanced Research Projects Agency under Agreement No. HR00112020021. Part of this work was completed while the second author was at Johns Hopkins University. Part of this work was completed while the third author was at Microsoft Research. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the United States Government or DARPA.

REFERENCES

[1] John M. Abowd, Gary L. Benedetto, Simson L. Garfinkel, Scot A. Dahl, Aref N. Dajani, Matthew Graham, Michael B. Hawes, Vishesh Karwa, Daniel Kifer, Hang Kim, Philip Leclerc, Ashwin Machanavajjhala, Jerome P. Reiter, Rolando Rodriguez, Ian M. Schmutte, William N. Sexton, Phyllis E. Singer, and Lars Vilhuber. 2020. The modernization of statistical disclosure limitation at the U.S. Census Bureau. https://www2.census.gov/adrm/CED/Papers/FY21/2020-08-AbowdBenedettoGarfinkelDahletal-The%20modernization%20of.pdf.

[2] Ruba Abu-Salma, Elissa M. Redmiles, Blase Ur, and Miranda Wei. 2018. Exploring User Mental Models of End-to-End Encrypted Communication Tools. In *Proceedings of the 8th USENIX Workshop on Free and Open Communications on the Internet (FOCI 18)*. USENIX Association, Baltimore, MD. https://www.usenix.org/conference/foci18/presentation/abu-salma

[3] Apple. 2018. Differential Privacy. https://www.apple.com/privacy/docs/Differential_Privacy_Overview.pdf.

[4] 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. https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/.

[5] Pablo J Barrio, Daniel G Goldstein, and Jake M Hofman. 2016. Improving comprehension of numbers in the news. In *Proceedings of the 2016 CHI: Conference on Human Factors In Computing Systems*. ACM, 2729–2739.

[6] Yoav Benjamini and Yosef Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 57, 1 (1995), 289–300.

[7] Sebastian Benthall, Seda Gürses, and Helen Nissenbaum. 2017. *Contextual integrity through the lens of computer science*. Now Publishers.

[8] Hugh Beyer and Karen Holtzblatt. 1999. Contextual design. *Interactions* 6, 1 (1999), 32–42.

[9] 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*. ACM, 3833–3837.

[10] Juan Pablo Carrascal, Christopher Riederer, Vijay Erramilli, Mauro Cherubini, and Rodrigo de Oliveira. 2013. Your browsing behavior for a Big Mac: Economics of personal information online. In *Proceedings of the 22nd International Conference on World Wide Web*. ACM, 189–200.

[11] Aloni Cohen and Kobbi Nissim. 2020. Towards formalizing the GDPR's notion of singling out. *Proceedings of the National Academy of Sciences* 117, 15 (2020), 8344–8352.

[12] Jamie Condliffe. 2019. The Week in Tech: Facebook's First Step Toward Treating Our Data Better. https://www.nytimes.com/2019/08/23/technology/big-tech-data.html.

[13] Lorrie Faith Cranor. 2006. What do they "indicate?" Evaluating security and privacy indicators. *Interactions* 13, 3 (2006), 45–47.

[14] Rachel Cummings and Deven Desai. 2018. The Role of Differential Privacy in GDPR Compliance. In *Proceedings of 2nd FATREC Workshop on Responsible Recommendation*. ACM.

[15] Inbal Dekel, Rachel Cummings, Ori Heffetz, and Katrina Ligett. 2021. Differential Privacy in Economics Experiments: A Public-Good-Game Application. Pre-print.

[16] Bolin Ding, Jana Kulkarni, and Sergey Yekhanin. 2017. Collecting Telemetry Data Privately. In *Advances in Neural Information Processing Systems 30 (NIPS '17)*. Advances in Neural Information Processing Systems, 3574–3583.

[17] Cynthia Dwork. 2008. Differential Privacy: A Survey of Results. In *Theory and Applications of Models of Computation*, Manindra Agrawal, Dingzhu Du, Zhenhua Duan, and Angsheng Li (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 1–19.

[18] Cynthia Dwork, Vitaly Feldman, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Aaron Roth. 2015. The reusable holdout: Preserving validity in adaptive data analysis. *Science* 349, 6248 (2015), 636–638.

[19] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. Calibrating Noise to Sensitivity in Private Data Analysis. In *TCC 2006 (LNCS, Vol. 3876)*, Shai Halevi and Tal Rabin (Eds.). Springer, Heidelberg, 265–284. https://doi.org/10.1007/11681878_14

[20] Cynthia Dwork, Moni Naor, Omer Reingold, Guy N. Rothblum, and Salil P. Vadhan. 2009. On the complexity of differentially private data release: efficient algorithms and hardness results. In *41st ACM STOC*, Michael Mitzenmacher (Ed.). ACM Press, 381–390. https://doi.org/10.1145/1536414.1536467

[21] Cynthia Dwork and Aaron Roth. 2014. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science* 9, 34 (2014), 211–407.

[22] Julia Brande Earp, Annie I Antón, Lynda Aiman-Smith, and William H Stufflebeam. 2005. Examining Internet privacy policies within the context of user privacy values. *IEEE Transactions on Engineering Management* 52, 2 (2005), 227–237.

[23] Serge Egelman, Raghudeep Kannavara, and Richard Chow. 2015. Is this thing on? Crowdsourcing privacy indicators for ubiquitous sensing platforms. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 1669–1678.

[24] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. 2014. RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response. In *ACM CCS 2014*, Gail-Joon Ahn, Moti Yung, and Ninghui Li (Eds.). ACM Press, 1054–1067. https://doi.org/10.1145/2660267.2660348

[25] Gerd Gigerenzer, Ralph Hertwig, Eva Van Den Broek, Barbara Fasolo, and Konstantinos V Katsikopoulos. 2005. "A 30% chance of rain tomorrow": How does the public understand probabilistic weather forecasts? *Risk Analysis: An International Journal* 25, 3 (2005), 623–629.

[26] Jens Hainmueller, Dominik Hangartner, and Teppei Yamamoto. 2015. Validating vignette and conjoint survey experiments against real-world behavior. *Proceedings of the National Academy of Sciences* 112, 8 (2015), 2395–2400.

[27] Moritz Hardt and Guy N. Rothblum. 2010. A Multiplicative Weights Mechanism for Privacy-Preserving Data Analysis. In *51st FOCS*. IEEE Computer Society Press, 61–70. https://doi.org/10.1109/FOCS.2010.85

[28] Eszter Hargittai and Yuli Patrick Hsieh. 2012. Succinct survey measures of web-use skills. *Social Science Computer Review* 30, 1 (2012), 95–107.

[29] Eszter Hargittai and Eden Litt. 2013. New strategies for employment? internet skills and online privacy practices during people's job search. *IEEE Security & Privacy* 11, 03 (may 2013), 38–45. https://doi.org/10.1109/MSP.2013.64

[30] Eszter Hargittai and Marina Micheli. 2019. Internet skills and why they matter. *Society and the internet: How networks of information and communication are changing our lives* (2019), 109.

[31] Michael Hawes. 2019. Title 13, Differential Privacy, and the 2020 Decennial Census. https://www2.census.gov/about/policies/2019-11-paper-differential-privacy.pdf.

[32] Ori Heffetz and Katrina Ligett. 2014. Privacy and data-based research. *Journal of Economic Perspectives* 28, 2 (2014), 75–98.

[33] Amaç Herdağdelen, Alex Dow, Bogdan State, Payman Mohassel, and Alex Pompe. 2020. New privacy-protected Facebook data for independent research on social media's impact on democracy. https://research.fb.com/blog/2020/06/protecting-privacy-in-facebook-mobility-data-during-the-covid-19-response/.

[34] Iulia Ion, Niharika Sachdeva, Ponnurangam Kumaraguru, and Srdjan Čapkun. 2011. Home is safer than the cloud! Privacy concerns for consumer cloud storage. In *Proceedings of the Seventh Symposium On Usable Privacy and Security (SOUPS)*. ACM, 1–20.

[35] Carlos Jensen and Colin Potts. 2004. Privacy policies as decision-making tools: an evaluation of online privacy notices. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 471–478.

[36] Noah Johnson and Joe Near. 2019. sql-differential-privacy. https://github.com/uber-archive/sql-differential-privacy.

[37] Ruogu Kang, Laura Dabbish, Nathaniel Fruchter, and Sara Kiesler. 2015. "My Data Just Goes Everywhere:" User mental models of the internet and implications for privacy and security. In *Eleventh Symposium On Usable Privacy and Security (SOUPS)*. ACM, 39–52.

[38] Gabriel Kaptchuk, Daniel G Goldstein, Eszter Hargittai, Jake Hofman, and Elissa M Redmiles. 2020. How good is good enough for COVID19 apps? The influence of benefits, accuracy, and privacy on willingness to adopt. *arXiv preprint arXiv:2005.04343* (2020).

[39] Shiva Prasad Kasiviswanathan, Homin Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. 2011. What Can We Learn Privately? *SIAM J. Comput.* 40, 3 (2011), 793–826. https://doi.org/10.1137/090756090

[40] Carmen Keller and Michael Siegrist. 2009. Effect of risk communication formats on risk perception depending on numeracy. *Medical Decision Making* 29, 4 (2009), 483–490.

[41] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W Reeder. 2009. A "nutrition label" for privacy. In *Proceedings of the Fifth Symposium On Usable Privacy and Security (SOUPS)*. ACM, 1–12.

[42] Katharina Krombholz, Karoline Busse, Katharina Pfeffer, Matthew Smith, and Emanuel von Zezschwitz. 2019. "If HTTPS Were Secure, I Wouldn't Need 2FA" - End User and Administrator Mental Models of HTTPS. In *2019 IEEE Symposium on Security and Privacy*. IEEE Computer Society Press, 246–263. https://doi.org/10.1109/SP.2019.00060

[43] Alexandra Mai, Katharina Pfeffer, Matthias Gusenbauer, Edgar Weippl, and Katharina Krombholz. 2020. User Mental Models of Cryptocurrency Systems - A Grounded Theory Approach. (Aug. 2020), 341–358. https://www.usenix.org/conference/soups2020/presentation/mai

[44] Aleecia M Mcdonald, Robert W Reeder, Patrick Gage Kelley, and Lorrie Faith Cranor. 2009. A comparative study of online privacy policies and formats. In *Proceedings of the International Symposium on Privacy Enhancing Technologies Symposium*. Springer, 37–55.

[45] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and inter-rater reliability in qualitative research: Norms and guidelines for CSCW and HCI practice. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–23.

[46] Frank McSherry and Kunal Talwar. 2007. Mechanism Design via Differential Privacy. In *48th FOCS*. IEEE Computer Society Press, 94–103. https://doi.org/10.1109/FOCS.2007.41

[47] Federico Morando, Raimondo Iemma, and Emilio Raiteri. 2014. Privacy evaluation: what empirical research on users' valuation of personal data tells us. *Internet Policy Review* 3, 2 (2014), 1–12.

[48] Vivian Genaro Motti and Kelly Caine. 2016. Towards a visual vocabulary for privacy concepts. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 60. SAGE Publications Sage CA: Los Angeles, CA, 1078–1082.

[49] Chaya Nayak. 2020. New privacy-protected Facebook data for independent research on social media's impact on democracy. https://research.fb.com/blog/2020/02/new-privacy-protected-facebook-data-for-independent-research-on-social-medias-impact-on-democracy/.

[50] Joe Near. 2018. Differential Privacy at Scale: Uber and Berkeley Collaboration. In *Enigma 2018 (Enigma 2018)*. USENIX Association, Santa Clara, CA. https://www.usenix.org/node/208168

[51] Helen Nissenbaum. 2004. Privacy as contextual integrity. *Wash. L. Rev.* 79 (2004), 119.

[52] Kobbi Nissim, Thomas Steinke, Alexandra Wood, Mark Bun, Marco Gaboardi, David R. O'Brien, and Salil Vadhan. 2017. Differential Privacy: A Primer for a Non-technical Audience. URL https://privacytools.seas.harvard.edu/files/privacytools/files/pedagogical-document-dp_0.pdf.

[53] Kobbi Nissim and Alexandra Wood. 2018. Is privacy privacy? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, 2128 (2018), 20170358.

[54] Daniel L Oberski and Frauke Kreuter. 2020. Differential privacy and social science: An urgent puzzle. *Harvard Data Science Review* 2, 1 (2020).

[55] Eyal Peer, Joachim Vosgerau, and Alessandro Acquisti. 2014. Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior research methods* 46, 4 (2014), 1023–1031.

[56] Elissa Redmiles. 2018. Net benefits: Digital inequities in social capital, privacy preservation, and digital parenting practices of US social media users. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 12. AAAI.

[57] Elissa M Redmiles, Yasemin Acar, Sascha Fahl, and Michelle L Mazurek. 2017. *A summary of survey methodology best practices for security and privacy researchers*. Technical Report.

[58] Elissa M Redmiles, Sean Kross, and Michelle L Mazurek. 2017. Where is the digital divide? A survey of security, privacy, and socioeconomics. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 931–936.

[59] Elissa M. Redmiles, Sean Kross, and Michelle L. Mazurek. 2019. How Well Do My Results Generalize? Comparing Security and Privacy Survey Results from MTurk, Web, and Telephone Samples. In *2019 IEEE Symposium on Security and Privacy*. IEEE Computer Society Press, 1326–1343. https://doi.org/10.1109/SP.2019.00014

[60] Christopher Riederer, Jake M Hofman, and Daniel G Goldstein. 2018. To put that in perspective: Generating analogies that make numbers easier to understand. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 1–10.

[61] Norman Sadeh, Alessandro Acquisti, Travis D Breaux, Lorrie Faith Cranor, Aleecia M McDonald, Joel R Reidenberg, Noah A Smith, Fei Liu, N Cameron Russell, Florian Schaub, et al. 2013. The Usable Privacy Policy Project: Combining Crowdsourcing. *Machine Learning and Natural Language Processing to Semi-Automatically Answer Those Privacy Questions Users Care About. Carnegie Mellon University Technical Report CMU-ISR-13-119* (2013), 1–24.

[62] Florian Schaub, Rebecca Balebako, and Lorrie Faith Cranor. 2017. Designing effective privacy notices and controls. *IEEE Internet Computing* (2017).

[63] Awanthika Senarath, Nalin AG Arachchilage, and Jill Slay. 2017. Designing Privacy for You: A Practical Approach for User-Centric Privacy. In *International Conference on Human Aspects of Information Security, Privacy, and Trust*. Springer, Springer, 739–752.

[64] Arun D Singh and John Paling. 1997. Informed consent: Putting risks into perspective. *Survey of Ophthalmology* 42, 1 (1997), 83–86.

[65] Paul Slovik. 1987. Perception of risk. *Science* 236, 4799 (1987), 280–285.

[66] Sarah Spiekermann and Lorrie Faith Cranor. 2008. Engineering privacy. *IEEE Transactions on Software Engineering* 35, 1 (2008), 67–82.

[67] Differential Privacy Team. 2017. Learning with Privacy at Scale. https://machinelearning.apple.com/research/learning-with-privacy-at-scale.

[68] Jessica Vitak, Yuting Liao, Priya Kumar, Michael Zimmer, and Katherine Kritikos. 2018. Privacy Attitudes and Data Valuation Among Fitness Tracker Users. In *Transforming Digital Worlds*, Gobinda Chowdhury, Julie McLeod, Val Gillet, and Peter Willett (Eds.). Springer International Publishing, Cham, 229–239.

[69] 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*. IEEE Computer Society Press, 392–410. https://doi.org/10.1109/SP40000.2020.00088

A COMPLETE SURVEY DESCRIPTIONS

For completeness, we include the language and flow of our two surveys below. Demographics for the two samples are included in Table 6.

A.1 Complete Description of Survey One

Respondents were randomized into either the salary scenario or the medical scenario, described below.

Salary Scenario: Imagine that you work in the banking industry. You are friends with a group of other people who work in banking companies in your city. One of your friends is part of a transparency initiative that is trying to publish general statistics about pay in the banking industry. As part of this initiative, they have asked everyone in the group to share their salaries and job titles using an online web form on the initiative's website.

Medical Scenario: Imagine that during your next doctor's visit, your primary care doctor informs you that they are part of a non-profit organization trying to push the boundaries of medical research. This non-profit is asking patients around the country to share their medical records, which will be used to help medical research on improving treatment options and patient care. Your doctor, with your permission, can facilitate the non-profit getting the information they need.

Questions: Answer options for each questions presented in <>. Text differences between the two scenarios presented in italics inside brackets.

- Which of the following would you want to know before deciding whether or not to share your *[salary/medical history]*? Select as many as apply.
  - *<[Whether your friend could learn your salary/Whether other doctors involved in the non-profit could learn your medical history]*
  - Whether a criminal or foreign government could steal your *[salary/medical history]*
  - Whether law enforcement could accesses your *[salary/medical history]* by obtaining a warrant
  - Whether data analyst at the initiative could see your *[salary/medical history]*
  - Whether graphs and charts created by the initiative could reveal your *[salary/medical history]*
  - Whether your *[salary/medical history]* could be shared with another organization
  - Other [free response]>

For each of the non-other options selected by the respondent, they were shown one of the follow three options, selected independently at random.

- *[leak entity]* will not learn your *[salary/medical history]*.
- *[leak entity]* might learn your *[salary/medical history]*. The chance this happens is about the same as the chance that your bank account will be compromised (accessed by a person who you did not intend to gain access to) as part of a data breach in the next year.
- *[leak entity]* might learn your *[salary/medical history]*. The chance this happens is higher than the chance that your bank account will be compromised (accessed by a person who you did not intend to gain access to) as part of a breach in the next year.

Finally, respondents were asked:

- Would you be willing to share your *[salary/medical record]* with the *[initiative/non-profit]*?
  <Yes, No, I'm not sure, Prefer not to answer>
  - *[If Yes]* Why would you be willing to share your *[salary/medical record]*?
  - *[If No]* Why would you not be willing to share your *[salary/medical record]*?
  - *[If I'm not sure]* Why are you unsure whether you would be willing to share your *[salary/medical record]* with the *[initiative/non-profit]*?

A.2 Complete Description of Survey Two

Respondents were randomized into either the salary scenario or the medical scenario, described below. In both scenarios, respondents were randomly shown a description of differential privacy, listed after the scenarios.

Salary Scenario: Imagine that you work in the banking industry. You are friends with a group of other people who work in banking companies in your city. One of your friends is part of a transparency initiative that is trying to publish general statistics about pay in the banking industry. As part of this initiative, they have asked everyone in the group to share their salaries and job titles using an online web form on the initiative's website. *[description from the list of differential privacy descriptions, shown below.]*

In this survey we are going to ask you a series of questions about a hypothetical scenario. Please do your best to imagine yourself in this scenario and answer the questions as if you were actually making the decisions about which you will be asked.

Medical Scenario: Imagine that during your next doctor's visit, your primary care doctor informs you that they are part of a non-profit organization trying to push the boundaries of medical research. This non-profit is asking patients around the country to share their medical records, which will be used to help medical research on improving treatment options and patient care. Your doctor, with your permission, can facilitate the non-profit getting the information they need. *[description from the list of differential privacy descriptions, shown below.]*

In this survey we are going to ask you a series of questions about a hypothetical scenario. Please do your best to imagine yourself in this scenario and answer the questions as if you were actually making the decisions about which you will be asked.

List of Differential Privacy Descriptions: Names of the description, shown in italics inside parenthesis, were not show to respondents.

- *(Control:) no additional text*
- *(Unsubstantial:)* To reduce the intrusion into personal privacy, your friend says they will use a technique called differential privacy. Differential privacy is the gold standard in data privacy and protection and is widely recognized as the strongest guarantee of privacy available.
- *(Techniques:)* To reduce the intrusion into personal privacy, your friend says they will use a technique called differential privacy. Differential Privacy injects statistical noise into collected data

	Survey One (n=1,216)			Survey Two (n=1,208)		
	Percent	**Mean**	**Stdev**	**Percent**	**Mean**	**Stdev**
Age	-	37.09	12.00	-	37.39	11.16
Woman	42.92%	-	-	40.64%	-	-
Man	56.08%	-	-	58.36%	-	-
Black	12.00%	-	-	11.92%	-	-
White	76.15%	-	-	76.98%	-	-
Hispanic	7.48%	-	-	7.12%	-	-
Asian	9.12%	-	-	7.53%	-	-
Native American	1.64%	-	-	1.32%	-	-
Edu. High school or less	9.04%	-	-	9.10%	-	-
Edu. Some College	23.60%	-	-	22.93%	-	-
Edu. Bachelor's or above	66.78%	-	-	67.54%	-	-
Income	-	US$61.6K	US$42.9K	-	US$61.5K	US$41.3K
Internet Skills (1-5)	-	2.19	.89	-	2.28	.86

Table 6: Survey demographics.

in a way that protects privacy without significantly changing conclusions.
- *(Enables:)* To reduce the intrusion into personal privacy, your friend says they will use a technique called differential privacy. Differential Privacy allows analysts to learn useful information from large amounts of data without compromising an individual's privacy.
- *(Trust:)* To reduce the intrusion into personal privacy, your friend says they will use a technique called differential privacy. Differential privacy is a novel, mathematical technique to preserve privacy which is used by companies like Apple and Uber.
- *(Risk:)* To reduce the intrusion into personal privacy, your friend says they will use a technique called differential privacy. Differential privacy protects a user's identity and the specifics of their data, meaning individuals incur almost no risk by joining the dataset.
- *(Technical:)* To reduce the intrusion into personal privacy, your friend says they will use a technique called differential privacy. Differential privacy ensures that the removal or addition of a single database item does not (substantially) affect the outcome of any analysis. It follows that no risk is incurred by joining the database, providing a mathematically rigorous means of coping with the fact that distributional information may be disclosive.

Questions: Answer options for each questions presented in <>. Text differences between the two scenarios presented in italics inside brackets. The names of the data leaks used in the main body of the text are shown in italics inside parenthesis and were not shown to users.

- Would you be willing to share your *[salary/medical record]* with the *[initiative/non-profit]*?
  <Yes, No, I'm not sure, Prefer not to answer>
  - *[If Yes]* Why would you be willing to share your *[salary/medical record]*?
  - *[If No]* Why would you not be willing to share your *[salary/medical record]*?
  - *[If I'm not sure]* Why are you unsure whether you would be willing to share your *[salary/medical record]* with the *[initiative/non-profit]*?

For each of the following statements, please indicate if you expect the following to be true or false if you share your salary and job title as part of this initiative.

- *(Organization:) [My friend will not be able to learn my salary and job title/The contents of my medical record will be stored only by my doctor's office, and will not be stored by the non-profit]*
  <Yes, No, I don't know>
- *(Hack:)* A criminal or foreign government that hacks the *[transparency initiative/non-profit]* could learn my *[salary and job title/medical history]*
  <Yes, No, I don't know>
- *(Law Enforcement:)* A law enforcement organization could access my *[salary and job title/medical history]* with a court order requesting this data from the *[transparency initiative/non-profit]*
  <Yes, No, I don't know>
- *(Data Analyst:)* A data analyst working *[on/for]* the *[salary transparency initiative/non-profit]* could learn my exact *[salary and job title/medical history]*
  <Yes, No, I don't know>
- *(Graphs:)* Graphs or informational charts created using information given to the *[salary transparency initiative/non-profit]* could reveal my *[salary and job title/medical history]*
  <Yes, No, I don't know>
- *(Share:)* Data that the *[salary transparency initiative/non-profit]* shares with other organizations doing *[salary/medical]* research could reveal my *[salary and job title/medical history]*
  <Yes, No, I don't know>