Journal of Privacy and Confidentiality
Vol. 9 (2) 2019

Submitted Aug 28, 2018
Published Oct 2019

DIFFERENTIAL PRIVACY IN PRACTICE: EXPOSE YOUR EPSILONS!

CYNTHIA DWORK, NITIN KOHLI, AND DEIRDRE MULLIGAN

349 Maxwell Dworkin, Harvard University, Cambridge, MA 02138
e-mail address: dwork@seas.harvard.edu

102 South Hall, UC Berkeley School of Information, Berkeley, CA 94720
e-mail address: nitin.kohli@berkeley.edu

102 South Hall, UC Berkeley School of Information, Berkeley, CA 94720
e-mail address: dmulligan@berkeley.edu

Abstract. Differential privacy is at a turning point. Implementations have been successfully leveraged in private industry, the public sector, and academia in a wide variety of applications, allowing scientists, engineers, and researchers to learn about populations of interest without specifically learning about individuals. Because differential privacy permits us to quantify cumulative privacy loss, these differentially private systems will, for the first time, enable the measurement and comparison of the total privacy loss incurred by these data-intensive activities. Appropriately leveraged, this could be a watershed moment for privacy.

Like other technologies and techniques that allow for a range of instantiations, implementation details matter. When meaningfully implemented, differential privacy supports deep data-driven insights with minimal worst-case privacy loss. When not meaningfully implemented, differential privacy delivers privacy mostly in name. Using differential privacy to maximize learning while providing a meaningful degree of privacy requires judicious choices with respect to the privacy parameter ϵ (among other factors). However, there is little understanding of what is the optimal value of ϵ for a given system or classes of systems, purposes, data, etc., or how to go about figuring it out.

To understand current differential privacy implementations and how organizations make these key choices in practice, we conducted interviews with differential privacy practitioners to learn from their experiences. We found no clear consensus on how to choose ϵ, nor agreement on how to approach this and other key implementation decisions. Given the importance of these details there is a need for shared learning amongst the differential privacy community. To serve these purposes, and foster competition, we propose the creation of the Epsilon Registry – a publicly available communal body of knowledge about differential privacy implementations that can be used by various stakeholders to drive the identification and adoption of judicious differentially private implementations.

Key words and phrases: Differential Privacy, Organizational Practices, Transparency, Interpretability.

JPrivacy Confidentiality

www.journalprivacyconfidentiality.org
DOI:10.29012/jpc.689

Introduction

Differential privacy is at a turning point, with the last three to five years witnessing implementation at a truly global scale. Examples of industrial use include: in the Google Chrome browser to identify vectors for malware [Erlingsson et al., 2014]; in Microsoft Windows' collection of usage and error statistics; and in all releases of Apple's macOS and iOS since 2016 to "identify things like the most popular emoji, the best QuickType suggestions, and energy consumption rates in Safari."[1] Open source implementations are available as well, including Google's TensorFlow Privacy for privacy-preserving machine learning.[2] Additionally, Microsoft and Harvard just announced a collaboration to build an open data differentially private platform.[3] Within the public sector, the US Census Bureau has deployed OnTheMap [Machanavajjhala et al., 2008], a privacy-preserving[4] web-based mapping and reporting application that shows not only where people work and where workers live, but also provides companion reports on age, earnings, industry distributions, and local workforce indicators. The disclosure avoidance subsystem of the 2020 US Census will employ differential privacy [Abowd, 2018].

Differential privacy allows us to quantify cumulative privacy loss as data are analyzed and re-analyzed, shared, and linked. These differentially private systems will, for the first time, allow us to measure and compare the total privacy loss due to these personal data-intensive activities. Appropriately leveraged, this could be a watershed moment for privacy in systems relying on personal information, as differential privacy is unique in enabling data subjects and other parties to assess the relative quality of (one component of) a firm's privacy practices prior to purchase or participation, permitting an informed decision.

As with other technologies and techniques that allow for a range of instantiations, the devil is in the details, and choices in the implementation of differential privacy influence the degree of protection against privacy loss in analogy to the manner in which the security parameter in a cryptosystem[5] determines whether the scheme will provide strong or meaningless protection of the messages. The privacy parameter, typically called ϵ ("epsilon"), provides a technical measure of privacy loss, with smaller ϵ corresponding to less privacy loss.

A compendium of results, colloquially known as the Fundamental Law of Information Recovery, tells us that overly accurate answers to too many questions can destroy any reasonable notion of privacy.[6] Experiments carried out by the US Census Bureau on summary statistics from the 2010 census are sobering. "Using only published contingency tables (summary statistics), we applied the database reconstruction theorem to reconstruct record-level images for all 308,745,538 persons enumerated in the 2010 census." Linking these reconstructed records to a commercial data set available in 2010 yielded confirmed

[1] See https://www.apple.com/privacy/ and https://images.apple.com/privacy/docs/Differential_Privacy_Overview.pdf.

[2] See https://medium.com/tensorflow/introducing-tensorflow-privacy-learning-with-differential-privacy-for-training-data-b143c5e801b6

[3] See https://www.iq.harvard.edu/news/harvards-iqss-announces-major-collaboration-microsoft and https://www.linkedin.com/pulse/microsoft-harvards-institute-quantitative-social-science-john-kahan

[4] This system that ensures a variant of differential privacy known as probabilistic differential privacy.

[5] For example, the number of bits in the decryption key.

[6] See Dinur and Nissim [2003] *et sequelae*; see Dwork et al. [2017] for a survey.

reidentification of 52 million of 308,745,538 persons, or 17% of the total population.[7] This imposes hard limits on what we can hope to achieve with any approach to private data analysis.

Using differential privacy to maximize learning while providing a meaningful degree of privacy requires judicious choice of the parameter ϵ (among other factors). But there is little understanding of what is a judicious value of ϵ for a given system or class of systems, purposes, data, *etc.*, or even how to go about figuring it out.

To understand current differential privacy implementations and how firms make these key choices – including but not limited to choice of ϵ – in practice, we conducted interviews with practitioners to learn from their implementation experiences. We found no clear consensus on how to choose ϵ or even how to approach this and other key implementation decisions; choices vary widely, resulting in systems that afford wildly different privacy guarantees; and, there is little collaboration, information sharing, or publishing to advance critical reflection on these key implementation issues. Given the importance of these implementation details there is a need for shared learning amongst the differential privacy community. In addition, disclosing information about firms' implementation choices is essential to drive meaningful differential privacy practices in the market.

To serve these purposes, we propose the creation of the Epsilon Registry – a publicly available communal body of knowledge about differential privacy implementations that can be used by various stakeholders to drive the identification and adoption of judicious differential privacy implementations. Firms using differential privacy would disclose the choice of ϵ, as well as several critical related policies and practices (we elaborate below). This information would be publicly available. The knowledge shared through the Epsilon Registry will advance privacy in two ways: it will support the identification of judicious parameter ϵ and other privacy preserving design choices and best practices among practitioners; and, by enabling stakeholders to compare the quality of privacy offered by various firms, create pressure on firms to reduce privacy losses while assuring utility gains.

1. Differential Privacy Implementations and the Quality of Privacy Practices

Differential privacy hides the presence or absence of any individual, or small group of individuals, in a dataset, in the sense that, for each individual, any conclusion reached from the analysis would be essentially as likely to have been reached, whether the given individual joined, or refrained from joining, the dataset. In this section we briefly define differential privacy, explore several of its important properties, and clarify its scope.

1.1. Defining Differential Privacy.

Differential privacy is a mathematical definition of privacy tailored to statistical data analysis. Mathematical definitions are distinct from the concepts and definitions discussed in privacy legal scholarship. Legal scholars attempt to separate distinct concepts of privacy, for example, informational, physical, decisional, and proprietary (see, *e.g.*, Allen [1998]). Formal legal texts, like statutes and case law, may define privacy affirmatively or in the breach, and often embody the principles in the Fair Information Practices (see Gellman [2017]).

[7] From remarks of John Maron Abowd, Chief Scientist, US Census Bureau, at the annual AAAS meeting, February 16, 2019, Washington, DC.

A mathematical definition is more specific still, and the definition of differential privacy defines a way of thinking about the problem of protecting a precise definition of privacy. It addresses interactions between a data analyst and a curator in possession of a dataset, requiring a specific relative bound on the probabilities of any given interaction on datasets that differ in the data of a single individual. In this way it prevents learning new information about an individual's data, but permits statistical learning about all individuals in the data set. Such statistical learning reveals information about classes of people, including information that individual people within the class might like to withhold. Differential privacy disentangles learning about a population as whole (declared not to be a privacy breach under this definition, like learning that smoking causes cancer) from learning idiosyncrasies of individual people that do not follow from information about the population as a whole (this smoker has cancer).[8] This is accomplished by introducing a controlled amount of randomness into the computation. This means that the output of a differentially private analysis depends not only on the data but also on the randomness, yielding (once the data set has been fixed), for each possible output, an associated probability that this output will be observed.

From the perspective of anyone whose fate depends on the outputs of the algorithm, we can say that, while it may be impossible to know the risk of a subjectively "bad" event, the decision to opt in or opt out of the data set will not significantly change the risk. Here "significantly" is controlled by the parameter ϵ, where a smaller ϵ means less change and hence better privacy. The change, which may be either an increase or a decrease, is by a multiplicative factor of at most e^{ϵ}.[9] When this bound e^{ϵ} is close to one (*i.e.*, ϵ is close to 0), anything that can be learned about an individual who has participated is almost equally likely to be learned about an individual who has *not* participated.

1.2. **The Opportunity for Meaningful Evaluation of Institutional Differential Privacy Practices.** Four features of differential privacy make it particularly valuable for the data-rich and data-driven environment.

First, the privacy loss in differential privacy implementations can be objectively *measured* via ϵ. This permits comparative privacy risk assessment between systems. While differential privacy does not draw neat lines between high-quality and low-quality implementations in all instances, it offers an opportunity to turn at least this component of an institution's privacy practices into a quality that can be ascertained by those outside the firm. This quantification allows for comparisons between institutions.

Second, the need to select ϵ creates an opportunity for reflecting on values. The ability to tune ϵ forces institutions to select a privacy-utility mix and document it, allowing a database controller to adjust her preference for privacy and utility. The controller may be the entity that owns or physically controls the database, or it may also be a regulator who holds logical control through a regulatory scheme. ϵ creates a placeholder for societal values in the loop of algorithmic decision making.

Third, it follows from the mathematical definition of differential privacy that being differentially private is independent of what a privacy adversary might or might not know,

[8]For a reconciliation of differential privacy and legal approaches to privacy see Chapter 4 in Groves et al. [2017]. Wood et al. [2018] provides an excellent differential privacy primer privacy for social scientists.

[9]When ϵ is much less than 1, this is approximately $1 + \epsilon$. When $\epsilon = 1$ this is approximately 2.71, and when $\epsilon > 1$ this grows very fast. For example, when $\epsilon = 3$ this is about 20, and when $\epsilon = 5$ it is about 148.4. And when $\epsilon = 15$ the value is greater than 3,269,017.

and to which other sources of information such an adversary may or may not have access, at any given time, even in the future. In other words, differential privacy is *future-proof* and *adversary-agnostic*.[10]

Finally, the mathematics of differential privacy allows the tracking – and the control – of cumulative privacy loss over multiple data uses.

1.3. **What Differential Privacy Doesn't Do.** Differential privacy is not a panacea, and in addition to understanding what it does provide, it is important to understand what it cannot provide, even with a tiny epsilon.

Differential privacy is the wrong tool to use to study outliers, as it hides their presence or absence. It is not the right tool for analyzing small datasets. Depending on the choice of epsilon, differential privacy may hide important differences in small populations or subpopulations of interest. While this may be construed as a limitation, it is actually a feature. Recall that differential privacy ensures that any conclusion reached from the analysis would be essentially as likely to have been reached with or without the data of any individual. Speaking informally, we expect statistical estimators run on large datasets to be more robust to the addition or deletion of a data point, as compared to the case with small datasets. In other words, adding or removing an individual from a small dataset is more likely to significantly change the value of the statistical estimator as compared to when the dataset is large. Despite lower accuracy, differential privacy is indeed working as intended – hiding the presence or absence of an individual over the outcomes of the analysis. In cases where small datasets need to be analyzed, sometimes legal constraints on data access, use, and sharing can support the right mix between discovery and privacy if the intended use warrants the confidentiality risks.[11]

Imagine a federal database of individually identifiable health information collected without consent for public health purposes, such as locating contagious individuals. Differential privacy does not address this nonconsensual collection, which intrudes on informational privacy.

Another potential concern is that the new knowledge derived from the privacy-preserving data analysis will be used to alter the information offered to individuals making an important health care decision (*e.g.*, a choice about reproduction). Differential privacy does not in any way constrain how the knowledge derived about the population as a whole is used to influence the treatment of specific individuals, whether their data were in the database or not.

1.4. **Reasoning about the Privacy Provided by Differential Privacy Implementations.** Although knowledge of ϵ is necessary to measure the privacy of a differentially private system, it is not sufficient. Numerous other design choices, as well as aspects of the data, affect the privacy provided by a differentially private system. Before undertaking our qualitative study to understand firm implementations, we considered differential privacy implementations in the abstract and identified a series of critical implementation choices. We discuss these implementation choices below and explain their potential impact on the privacy provided by a differentially private implementation.

[10]This immunizes differential privacy from the leveraging of external data, as in the famous de-anonymization of members of the Netflix Prize dataset [Narayanan and Shmatikov, 2008].

[11]See https://github.com/frankmcsherry/blog/blob/master/posts/2016-02-03.md.

1.4.1. *The Importance of ϵ for Assessing Privacy Quality.* The definition of differential privacy does not say how much privacy we need – as a general matter or in any given context. Not all data are of equal sensitivity, and, importantly, not all data usage is of equal value. We may be willing to provide weaker protection where we are pursuing a cure for cancer and/or when differential privacy is used in concert with employment contracts, ethics codes, public laws or other constraints. The parameterization of differential privacy allows different normative decisions about how much privacy to ensure. It creates a particular space for this significant social choice which ultimately controls the values of the algorithmic outputs: it holds out the possibility of putting society directly in the loop [Rahwan, 2018].

1.4.2. *Limitations on Total Privacy Loss.* Recall that the Fundamental of Information Recovery tells us that overly accurate answers to too many questions can destroy any reasonable notion of privacy. To protect against this, differential privacy in practice requires a privacy *budget* to limit harm. This budget is the (declared) maximum acceptable privacy loss allowed before no more queries are permitted. Once this budget has been exhausted, the dataset is retired, never to be queried again.

For concreteness, let's consider an example where we have a differentially private mechanism with a privacy budget of 2. There are multiple ways to fully utilize the budget. One possible way is to perform two computations, each with ϵ set to 1. Another way is to run three computations: the first with $\epsilon = 1$, and the remaining two with $\epsilon = \frac{1}{2}$. In both cases, the total privacy loss is bounded by 2, but the budget was "spent" in different ways. In the first example, the same amount of privacy was lost in each computation. In the second example, the second and third computations had half as much privacy loss compared to the first computation, but at the cost of utility.

1.4.3. *At What Point is Differential Privacy Applied?* Two differentially private systems can provide radically different privacy protection depending upon when differential privacy is introduced. Small values of ϵ always yield good privacy protection (see Section 1.5.2 below), so to get a feel for the difference it is easier to think about very large values of ϵ. Suppose the goal is to find out how many people in the database floss regularly. The following two algorithms have infinite privacy loss: (1) reveal the name and flossing practice for each individual in the database; (2) reveal only the statistic that was requested.

We can create a pair of contrived differentially private algorithms for the flossing problem with very large privacy loss that will behave analgously to the pairs in the infinite loss example. These correspond, respectively, to applying (very bad) differential privacy to the raw data and computing on the result and computing the result and then applying differential privacy. This also illustrates that the privacy loss does not by itself tell the whole story.

1.4.4. *Privacy Over What.* A movie recommendation system that protects the privacy of the viewer at the single-movie level – meaning that, for any given movie the system will provide ϵ-differential privacy for individual movie-watching events – is not providing nearly so much protection as a recommendation system that provides differential privacy for the entire movie watching history of the user. For example, if the viewer has a penchant for pornography, a per-movie privacy guarantee does not hide this viewing trend, while a per-complete history

privacy guarantee does so. This question is the *granularity* at which differential privacy is applied.

1.5. **The Meaning of Epsilon.** Uncertainty in choosing ϵ has two sources: uncertainty about social optimality; and uncertainty of how large epsilons function.

1.5.1. *Uncertainty about Optimality.* Pressure for good differential privacy implementations is undermined by the lack of clarity of what is the right degree of privacy loss in a given context. First, a method for extracting the best possible utility for any given value of ϵ may simply not be known, both for a given analytical task in general or for a very specific type of data. This can be addressed, over time, with technical research. Second, we lack a formula for determining, for a given privacy-utility tradeoff, what is the judicious choice of ϵ. This is especially true given that, as in the flossing example, not all large ϵ are alike.

Thus, while we can cap the privacy loss of a given algorithm, if we do not know how small an ϵ is possible – how much privacy can be offered – consistent with a given analytical utility (the optimal privacy-utility mix), or if we don't know a fair privacy price to pay for what we learn, the measurement has less meaning.

1.5.2. *Unifying the Meanings of Small Epsilons.* While all small ϵ are alike,[12] each large ϵ is large after its own fashion,[13] making it difficult to reason about them.

Imagine a world inhabited by two types of beings: ghosts and humans. Both types of beings behave the same, interact with others in the same way, blog, study, work, laugh, love, cry, reproduce, become ill, recover, and age in the same fashion. The only difference is that ghosts have no records in statistical databases, while humans do. The goal of the privacy adversary is to determine whether a given 50 year old, the "target", is a ghost or a human. Indeed, the adversary is given all 50 years to do so. The adversary does not need to remain passive, for example, she can organize clinical trials and enroll patients of her choice, she can create humans to populate databases, effectively creating the worst-case (for privacy) databases discussed above, she can expose the target to chemicals at age 25 and again at 35, and so on. She can know everything about the target that could possibly be entered into any database. She can know which databases the target would be in, were the target human, so that the only possible privacy breach left to worry about is the target's human/ghost bit. In this case each guarantee of a bound of ϵ on privacy loss is comparable: the specific algorithm and type of data is irrelevant.

When ϵ is small, whether we are talking about one computation or the cumulative privacy loss over multiple computations on the same database or over multiple databases, the human/ghost bit stays hidden, no matter how much or how little additional information the adversary has about the target and about the rest of the world. In this case small ϵ ensures that the database leaks essentially nothing about the target (beyond what would be revealed about the target given the result of the computation were the target to have opted out of the dataset).

As we have seen in the flossing example, this unification fails when ϵ is large. Mathematically, it is because working with a large ϵ simply fails to hide the human/ghost bit, so

[12] Recall that $e^\epsilon \approx 1 + \epsilon$ for $\epsilon << 1$

[13] This is based on an analogy to the opening sentence of Anna Karenina:"All happy families are alike, but an unhappy family is unhappy after its own fashion."

we cannot make any deduction from the assumption that this bit is hidden (as we just did in the case of small ϵ).

Thus, sound reasoning about the semantics of differential privacy, including ensuring that we are legitimately comparing apples to apples, when ϵ is small, becomes unsound when ϵ is large.

2. Case Studies - Differential Privacy in Practice

In the previous section, we described the fundamental characteristics of differential privacy. It is measurable, future-proof, adversary agnostic, and composes automatically. Furthermore, given its precise mathematical formulation, differential privacy allows for formal reasoning about privacy loss from statistical learning. All things equal, the quality of privacy provided by such mechanisms hinges on the value of ϵ, for which, other than the mantra that "small ϵs are happy epsilons," there is no hard-and-fast rule for its determination in general. But all things are rarely equal, and there are many properties of datasets and many implementation decisions that alter what privacy is provided by a differentially private system even with the same ϵ. Given this disparity, we set out to learn about existing implementations, documenting the choices and to some extent the decision making process.

2.1. **Research Methodology.**

To understand the process of choosing ϵ in practice, we conducted an empirical study of institutions known to be utilizing some form of differential privacy in 2016. Since our goal was to learn how organizations addressed implementation issues around differential privacy, we chose our unit of analysis to be the organizations themselves. Within each organization, we interviewed individuals with detailed knowledge of the specific design and institutional choices made during the implementation. In some cases, we interviewed multiple employees from the same organization to paint a complete picture of the differentially private ecosystem housed by their organization. This yielded 11 individuals from 7 organizations, which at the time of the interviews was a census of all organizations we knew to be utilizing differential privacy.

These organizations themselves were from diverse sectors: three were technology companies, and the remaining institutions were involved in government, telecommunications, energy, and academic research. The interviewees themselves held heterogeneous titles within their organizations: engineer, scientist, researcher, and director. Despite varying job titles, all of the individuals selected were deeply involved, in some capacity, with the implementation of their organizations' differential privacy system.[14]

As we will see in the following section, the same questions and challenges emerged across all organizations. While our interviews were conducted in 2016, and there are additional organizational implementations today, the implementation choices we identified as significant that were confirmed as challenging by through our interviews have not changed. There is no greater clarity today about how to choose ϵ or other parameters, and discussions with individuals about more recent implementations of differentially private systems suggest that they are struggling with the same set of questions.

[14]It is worth noting that while we were interviewing these specific individuals for specialized knowledge and experience, we were doing so only to learn about the organizational decisions about differentially private implementations. We were not interested in studying the behavior of the individuals. Given this scope, we did not go through an Institutional Review Board prior to conducting these interviews.

2.2. **Interview Format & Topics Probed.** The interviews were conducted by telephone from July 2016 until October 2016, each lasting approximately two hours. The questions selected were generated by the authors' knowledge of differential privacy, institutional privacy practices, and the abstract thought exercise discussed above in Section 1. These questions explored key design choices that institutions made during differential privacy implementations and the institutional impetuses for employing differential privacy, as well as any barriers encountered along the way.

Careful consideration was taken to present these questions in a broad manner, so as to allow interviewees to fully articulate their experiences and to avoid response bias and unintentional priming. A complete list of the questions can be found in Appendix B. At a high level, the interviews followed a protocol that focused on five categories:

(1) *Impetus and Barriers to Differential Privacy Adoption*: These questions explored the organizations' needs for privacy protection, and what spurred them to choose differential privacy at all, and to choose it over alternative privacy protection schemes. This section also explores barriers to adoption within the organizations.
(2) *Data Characteristics & Avenues for Loss*: These questions focused on the qualities of the raw data that differential privacy was to be used over. For example:
   - What kind of data was differential privacy being used on?
   - What were the potential avenues for privacy loss for the particular data and task in question?
(3) *Granularity of Protection*: This addresses the unit of analysis for the implementation. As such, this question is concerned with understanding what privacy protection is being offered over. For example, is differentially privacy used over an individual's entire history or only over events individually? What other options were considered, and why was this level chosen?
(4) *Limitations on Privacy Loss*: These questions were concerned with any organizational limits placed on privacy loss. If there were limits, to what time period did the limits apply? For example, one might constrain privacy loss to $\epsilon = 1$ per day, or per month.
(5) *Algorithm Specifics*: These questions were interested in the parameters and specifications of the differentially private algorithm. For example:
   - What variant of differential privacy is being used? Was it ϵ-differential privacy, the above-mentioned approximate differential privacy (also known as (ϵ, δ)-differential privacy), or another flavor?
   - Once the particular type of differential privacy was chosen, how were the parameters selected?
   - What factored into this consideration? And who from the organization participated in determining these values?

It is worth noting that these topics are truly of importance to holistically understand a differentially private implementation. As discussed in Section 1, we know that judging a system based solely on ϵ is not sufficient for a thorough assessment of privacy quality. Any inquiry into the nature of a differentially private implementation must examine the system as a whole: data inputs to the system, as well as all the protections to be achieved.

3. Results of the Case Studies

The interviews yielded critical insights about how differential privacy is used in practice. We found that organizations did in fact grapple with the implementation issues we had identified,

and had varying approaches to overcome, and in some cases ignore, these problems. In the sections that follow, we describe these findings.

3.1. **Impetus for and Barriers to Differential Privacy.** We found the driving force for differential privacy varied across organizations. Several interviewees noted their organization "*had taken a stance on privacy paramount [...] To adhere to that standard we needed a rigorously well-founded notion. [Differential privacy] was one of them because it has both privacy and, in many cases, strong utility guarantees.*" As such, the adoption of differential privacy was not an enormous leap.

Others had a different experience. They noted that the desire to use differential privacy did not come from the legal or privacy departments – rather, it was championed by engineering, science, and research divisions.

> *"We wanted something that gave good guarantees. We were pleased when we came up with something that gave us what we wanted. [...] The privacy team had a suspicion it would not work out. It sounded too much like magic."*

Two of the main barriers reported during the interviews were related to communication issues between engineers and other members of the organization, as well as skepticism about the theoretical promises of differential privacy.

> *"[Coming from research,] I had communication gap with product org. [...] For me a solution was a theorem [...] In a lot of the cases the people I worked with were not theoreticians or exposed to math. They were unexcited by theorems. So I had to write implementations to show things were converging to something demo-able."*

Fascinatingly, this is the opposite direction of how research goes. Typically researchers start with small prototypes and generalize to a larger class of examples via theorems. Once a theorem is established, all examples must fall under its purview. To bridge this gap, several interviewees found simulation and examples to be an effective way to gain buy-in and communicating with other members of their organization. Two interviewees noted:

> *"The thing with the simulation is that it is really easy and quick to do. We used simulations and [slider bars] that people could drag and drop. [This] convinced people it was worth pursuing as long as our sample sizes were large."*
>
> *"[Simulation was] important for whomever we communicated with. [Differential privacy] is a slightly complicated technology. Partly it seems to be black box. Theorems [were] not helpful. [But, folks can see] that in simulation we are converging to that number."*

The success of these methods suggests that simulations are a powerful way for non-experts to understand technical design. Where theorems failed, examples and experiential learning flourished. These solutions succeeded in building confidence in differential privacy as a technique by demonstrating there was still utility in the results that were being delivered and it was useful for solving product and business problems.

3.2. **Data Characteristics & Avenues for Privacy Loss.** When considering data characteristics and avenues for privacy loss, we found that our interviewees had the foresight to think through the ways in which their domain-specific information could be leaked via usage. One respondent, who works specifically with web browsing, noted that the avenues for privacy loss for his use cases were "*correlation between sites. The fact that sites do cluster and people do go to sites that share their language or interests. There is obviously this loss.*" Alternatively, a different respondent who works with operating systems noted that "*as you bring in more and more types of data, even as you try to make it impossible to link, now you can use correlation through metadata.*"

In both of these examples, the avenues for privacy loss involved domain-specific attributes. Context mattered - avenues for privacy loss were not consistent across all differentially private systems. In spite of the fact that our interviewees came from diverse fields, all avenues of privacy loss followed the general theme of re-identification via correlations (rather than inferring a particular property of an individual via indirect analysis and computation).

3.3. **Granularity of Protection.** When considering using differential privacy for protection, a natural question arises - differential privacy over what? Does the implementation seek to protect event level data, individual data, or something else? Generally, organizations were aware of the granularity of data they were trying to protect. However, we found that they were thinking about granularity in different ways.

Methods for determining the appropriate level of granularity differed, as they were tied to the specific data and use cases in question. The granularity used in one implementation may be completely inappropriate in another. For example, a researcher working on browser usage noted that event level information, such as the domain name, was critical for their organization's use, while the specifics of any particular user's behavior was not. When asked why they chose this unit of analysis, the researcher noted,

> *"Differential privacy is applied to the domain name [because that] is what is being protected."*

Others were thinking of granularity in terms of an individual person only. For example, when asked why the choice was made at the user-level and not the event level, a different interviewee noted,

> *"Definitely, you could do at the data record level, but it does not match the intuition that individuals want to be protected. User-level is much better fit for that, even though it is harder to achieve. We felt that was important to achieve."*

3.4. **Limitations on Privacy Loss.** Determining an appropriate privacy budget is context specific. When probing into this topic, we found results that differed by orders of magnitude. One interviewee noted, "*[We are] thinking about ϵ on the order of 1,2, and 4 [over a year]*" while another interviewee, from a different organization, stated the "*overall ϵ budget is 0.1.*"

On the topic of privacy budgets, our interviewees noted the difficulty in explaining the notion of a privacy budget. Three different interviewees from three different organizations noted that,

> *"[There was] lots of educating [...] They don't understand why I am so insistent on quantifying total budget."*

> *"[Randomized Response] is relatively easy to explain. Cumulative damage to privacy is sometimes confusing."*
>
> *"The non-technical people have a challenge where we tell them that the privacy budget is exhausted, and you can't answer any more queries. Doing this fully properly is very challenging. What more can I say about this as a scientist. I have not made much progress."*

This communication issue stems from a knowledge gap between individuals familiar with differential privacy and those less comfortable with the underlying mathematical model. While our interviewees tried to educate their peers about the importance of the privacy budget, the gap in knowledge presented a formidable barrier in this aspect.

3.5. **Algorithm Specifics.** With respect to techniques for determining ϵ, we found a wide range of methods being utilized. The approaches ranged from fairly sophisticated to, more-or-less, a random choice.

On the more sophisticated end of the spectrum, there were many diverse methods for determining an appropriate set of parameters. The first was experimental in nature. "*There are the 3 knobs we've played with. We tested with simulation.*" Simulation was used to find a value of ϵ that provided sufficient utility to meet an institution's specific product and business goals.

Others used a value of ϵ that was rooted in a previous differential privacy implementation. "*We took [the prior implementation] as a starting point. We anchored.*" This involved computing a new value of ϵ to meet the same utility produced from a prior differential privacy implementation.

Another approach was to build a threat model and work backwards to deduce necessary values for protection. The threat model was constructed by along the following lines.

> *"Given how often different factors probably change in the real world and empirical distribution, what are the odds you can identify them and breach their privacy? [There is a] risk of repeated reports and collection. We chose our ϵ based on that, specifically, what do we need to give us the protections we need. Then [we] looked at the accuracy. Then dialed our sampling back. Even with the protection we wanted, we had more than enough accuracy and could down-sample."*

On the opposite end of the spectrum, there were practitioners who admitted that the choice of ϵ was completely arbitrary without much consideration. "*Currently overall ϵ budget is* 0.1 *and delta is* 10^{-5}*. It is an arbitrary choice.*" This approach adheres to the "small ϵ are happy epsilons" mantra, without other consideration about whether this is truly appropriate for the use-case in question.

It is worth noting that almost all of these implementations err on the side of utility over privacy. The only approach that put privacy protection ahead of utility was the threat-model construction. This result is not too surprising, as organizational incentives dominate ainternal decision-making. In spite of this widespread preference of utility over privacy, there was no general agreement on how to choose ϵ.

3.6. **Takeaways from the Interviews.** We found that institutions faced similar challenges during the implementation of differential privacy. Some of the challenges were design based in nature, while others were institutionally based due to necessary interaction with individuals from different parts of the organization and with different levels of computational knowledge. While individuals devised solutions to address problems that arose during implementation, such as simulations for experiential learning, there was no consistency in approaches across institutions. While our subjects struggled with a similar set of choices, they made different decisions, based on different processes and considerations.

The range of practices and choices we identified suggests a need for shared learning amongst the differential privacy community. Without this shared learning, the nuances of differential privacy that arise in practice, and the solutions that follow, are unlikely to be shared across the privacy landscape.

The variation in implementations also reveals that a key benefit of differential privacy to the market and regulatory landscape – its support for objective and testable statements about privacy – remains latent. Policymakers, regulators, business partners, customers, and the public cannot assess the absolute or relative benefit of a given differential privacy implementation absent information about key choices made during implementation.

Thus, both practitioners and intermediaries in public oversight roles would benefit from additional information about differential privacy implementation specifics.

4. The Epsilon Registry

We do not know what parameter ϵ is right for any given differentially private analysis, and we do know that the answer can vary tremendously based on attributes of the dataset and the policies and practices that constrain those who query it. This flexibility is a virtue and a risk. It allows for context sensitive deployments. However, as we have seen, when ϵ is large it can also allow for a form of privacy theatre – the technique is used, but so weakly implemented that it offers little to no protection.

The discoverability of ϵ and other implementation details allows for distinguishing relatively good (protection) from bad use of differential privacy. Where implementation details are not disclosed, the protection offered by differential privacy remains as unknown as the privacy protection offered by other firm privacy practices (from encryption implementations, to privacy impact assessments, to records management policies and procedures). The inability to distinguish between variations in the quality of differential privacy implementations will undermine the privacy protective potential of differential privacy.

Broad knowledge of parameter ϵ choices across firms is important for two additional reasons. First, the judicious parameter ϵ for particular contexts, research goals, or datasets is not self-evident, but will be developed through trial and error. Public knowledge of parameter ϵ choices builds a knowledge base to support shared learning. Second, the privacy ramifications of parameter ϵ are enormous. In some instances, for example where parameter $\epsilon = \infty$, it may be the single point of failure. The hidden nature of this exceedingly material term will burden regulators tasked with protecting privacy. Given the limited additional cost to firms of disclosing parameter ϵ (it is information the firm must have and its release does not harm privacy of the data) and the important role it can play in assuring the availability of high quality implementations of differential privacy, disclosure is desirable.

A first question is whether firms will voluntarily disclose their chosen values for ϵ. The lack of specificity about the technologies and practices used by companies to protect the

security and privacy of personal information today suggests that the market alone won't drive such disclosures by companies. With some notable exceptions – for example the use of the lock in browsers to indicate encrypted data transfers – entities make few specific disclosures about methods of protecting personal information. The lack of details about technologies, techniques and practices may reflect a general belief that limiting information about defenses makes it costlier to mount attacks. It may be that firms have concluded that technical details are generally of little use to buyers who lack the sophistication to interpret them. It may be that greater specificity about the means of protection creates work for those responsible for maintaining the accuracy of policy statements to the public. Finally, the potential legal liability created by inconsistencies between public statements and actual implementations may push firms toward vague and general statements. Regardless of the root causes, firms rarely provide operational details, and we have little reason to suspect that they will behave differently with differential privacy. It may be that firms will not mention the term differential privacy at all. As with the term "privacy policy" which is today used to describe policies and practices that provide little in the way of privacy protection – as well as those that do – without ϵ disclosures the range in quality of differential privacy creates an opportunity for marketplace confusion.[15] ϵ disclosure is thus essential to privacy's protection, and unlikely to occur absent additional incentives.

An information disclosure law is well-suited to address our concerns [Magat et al., 1992, Sunstein, 1998]. Where, as here, firms have little incentive to disclose voluntarily, it is appropriate to use the coercive power of the state to make information available and thereby reduce the transaction costs that frustrate rational behavior by others [Sunstein, 1992]. By reducing the cost of acquiring relevant information, the law increases the chance that all affected individuals can make decisions that reflect their interests and values. Some such laws are aimed at addressing information asymmetries in marketplace transactions, such as food labeling about calories and nutritional makeup or SEC disclosures to enable better investment decisions through better information about corporate assets and liabilities. Others reflect more directly the political call that spurred their adoption – the "right to know"[16] – focusing on opening up firm practices to greater scrutiny where their effects may be non-obvious, of keen interest to the public, and yet unlikely to be captured through shifts in individuals' market place decisions. Our proposal is of the latter variety, designed to empower a range of stakeholders who can drive the adoption of privacy protective differential privacy systems, not just direct parties to a transaction.

Below we propose the adoption of an Epsilon Registry, into which firms must deposit information about differential privacy implementations. The Registry will support shared learning, allowing organizations to learn from others' choices. Additionally, the Registry serves as a mechanism for oversight by providing transparency into the set of practices and ϵ choices used by institutions. However, our aim does not stop there. This disclosure also enables a range of stakeholders – researchers and practitioners seeking to identify the most

[15] See Turow et al. [2007]. Also, Lee et al. [2005] found that the inclusion of a privacy policy increased the probability of an online purchase – increased the perceived trustworthiness of a web site – despite the relative ease with which bad actors could use this signal. In their assessment consumers were overly reliant on a weak signal. They didn't take into account the potential enforcement risk of false signaling, with respect to privacy or other factors they consider, which may be a somewhat significant deterrent given enforcement activities around deceptive privacy policies.

[16] Freedom of Information Acts (FOIA), sunshine acts focused on public monitoring of governmental decisions. For example, 5 U.S.C.A. Section 552b, known as the Government in the Sunshine Act, requires in part that all meetings of government agencies be made public.

judicious algorithms, regulators seeking to limit privacy theatre – who care about privacy practices in the marketplace, to progress the state of the art and share the benefits across the privacy landscape.

5. Registry Contents

We have argued that knowledge of ϵ is necessary to assess the quality of a differential privacy implementation, and noted that in some cases it may be insufficient on its own. As we have already observed, the significance of a large epsilon is ambiguous, and in all cases other information about the data set and governance practices are necessary for a meaningful evaluation. Thus, our Epsilon Registry requires public and private institutions, for each dataset, to disclose aspects of their differential privacy implementations, including:

5.1. **Paths of privacy loss.** What are the paths of privacy loss? What are the uses of the data during which information regarding one person can affect the experience of, or otherwise be exposed to, another person? For example, in audience size estimation for targeted advertising, by definition the counts of the number of users satisfying certain criteria are released to the advertiser (who may be a firm, researcher, or stalker).[17] The billing system is another pathway, as exploited by Korolova [2010].[18] In movie recommendation systems, the viewing habits of one user affect the recommendations made to a different user. The whole point of a recommendation system, be it in advertising, search engine results, or entertainment suggestions, is to leverage information of many people in order to predict response to candidate recommendations. With skill – and data – this can be performed in a differentially private way. Done clumsily, we get privacy breaches [Calandrino et al., 2011], as recorded in the Canon of Counterexamples (Appendix A).

5.2. **Granularity.** Thinking of differential privacy as providing protection about whether a specific datum, or piece of information, is or is not in the database, the granularity of protection describes the "type" of the datum protected. For example, a patient's record in a hospital database contains many data fields, or attributes. What is the granularity of the in/out protection: Is each attribute of an individual protected at level ϵ, leading to a loss of many multiples of ϵ when the database contains many attributes, or is the individual's entire medical history given ϵ-differential privacy in toto? In the movie recommendation system example, is the datum a single movie, or is it an entire movie-watching history? We have already discussed how a movie recommendation system providing "event-level" privacy of individual movie-watching events at a given level ϵ of differential privacy provides much less protection than a system offering ϵ differential privacy for the entire movie-watching history of the user (known as "user-level" privacy), so the granularity tells a lot about the overall quality of the protection afforded.[19] Information relating to database schemas and the types of queries allowed should be included as well, as these would strengthen transparency about

[17] Even if multi-way marginals are protected by rounding, complete behavioral and demographic profiles consisting of hundreds of attributes can be reconstructed (Mironov and Talwar, private communication, 2014).

[18] See Appendix A for more details

[19] Event-level and user-level protections do not always diverge to such a great extent. For a website, such as a flu self-assessment site, that is typically visited only a handful of times by any given individual, user-level privacy and event-level privacy are close for most individuals.

potential privacy risks. This information is useful for data subjects, regulators, Institutional Review Boards (IRBs), and privacy researchers.

5.3. **Epsilon per datum.** The pathways describe the avenues for privacy leakage, that is, the uses of the data that leak information; the granularity describes the unit of information that is protected. For every pathway, what is the granularity of protection and what is ϵ for each datum?

5.4. **Burn rate.** Many analyses, such as trend/popularity monitoring, are carried out continually. Data involved in these analyses may maintain their importance or may be down-weighted as they age. Other analyses, such as building click predictors for advertisements, are computationally intensive, "touching" each training datum many times; moreover the "signal" in the data may change hour to hour, as well as exhibiting different patterns on different days of the week, so the analysis is carried out frequently. Freeway traffic patterns may have similar cyclic behavior, but we can exploit its relative predictability with a different kind of analysis ensuring that privacy loss only occurs when the pattern is broken. The burn rate measures the cumulative privacy loss per user per unit time. Audience size estimation is performed on demand (this amounts to a set of counting queries). What is the rate of burn permitted in these sorts of activities, or is there no limit?

5.5. **Privacy loss allowed before retirement.** Are data "retired" after they have reached a certain degree of exposure (e.g., worst-case or expected cumulative privacy loss?) What is the threshold for retirement?

5.6. **Variant of differential privacy used.** What variant of differential privacy is in use? The literature contains at least three variants: "pure", or $(\epsilon, 0)$-differential privacy; "approximate", or (ϵ, δ)-differential privacy; and "probabilistic" differential privacy (used in OnTheMap). A few more, such as "concentrated" differential privacy (better accuracy and same behavior under composition as approximate differential privacy), are under investigation [Dwork and Rothblum, 2016, Bun and Steinke, 2016, Mironov, 2017].[20]

5.7. **Justifying implementation choices.** How were (5.1)-(5.6) chosen? What were the assumptions, modelling decisions, thresholds, and subjective decisions made in determining the implementation choices above? Why is the approach a thorough test of the stated assumptions? Was the process validated and verified? If so, how? This section acts as a signal of a firm's commitment to protecting privacy. In particular, this is where practitioners can share other relevant information about their implementations, as well as searching for what other practitioners have done. It displays what specific information went into the choice of ϵ.

Finally, research studies should publish total privacy costs and the ϵs for released statistics. Research is often exploratory; an examination of a dataset may have many false starts and dead ends that, in the current environment of scientific publishing, never see the light of day. Privacy loss accumulates during the exploratory phase. These losses should be

[20]All mentioned variations cope with arbitrary auxiliary information and are future-proof.

published. Certain statistics will ultimately be published; these published statistics should be accompanied by the ϵs with which they are computed.[21]

6. Disclosing and Justifying Implementation Choices

A system's privacy properties are important not only from a technical perspective, but also from a social one. As previously noted, unless pertinent information is made available, assessment and comparison of differentially private systems is infeasible.

Disclosure of the first six Registry contents provide insight into a system's technical specifications. The last component, justifying implementation choices, sheds light on a separate aspect – the considerations and choices made during the implementation process. This component documents why an organization believes its implementation choices are appropriate, and captures how the firm values privacy in light of the utility it costs.

An explanation of a firm's implementation choices signals the credibility and thoughtfulness of the privacy quality offered. By explicitly requiring explanations of these choices, the last component incentivizes against the case where a firm makes arbitrary choices without consideration of the implications, such as choosing an ϵ that seems "small enough" without much forethought.

For practitioners, the last Registry component aids in shared learning. Donoho stated [Stodden, 2012],

> *An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures.*

Along the same vein, disclosure of the first six Registry components alone is an advertisement of privacy quality. They state technical specifications of the system, but do not reveal how or why they were chosen. They share the decisions made, but not the rationale or the process.

The last tenet of the Registry provides insight into a firm's implementation methodologies and justifications. Similar to the best practice of disclosing research methods and justifications in the reproducible research movement to ensure reliable knowledge about computational and data-driven sciences [Stodden and Miguez, 2013], this tenet supports robust learning and knowledge transfer. Having firms disclose the high-level descriptions of the methodologies used, with the parameter details and the particular decision made, is useful in evaluating and learning about differentially private systems. By removing the opaqueness around the system and the design choices, the Registry allows for all practitioners to know not only the approaches and considerations used by their peers, but also what concerns and nuances are being considered throughout the entire process. Additionally, these explanations

[21] One might mistakenly think that, since the computations in the false starts are not published they cause no privacy loss. The error is that the results of the computations in the false starts affect even the choice of statistics that ultimately are computed and published. As an extreme contrived example, imagine that if the data of the Speaker of the House are included in a dataset, then the researcher publishes statistic S, and otherwise the researcher publishes statistic T. Even if differential privacy is used for the computation of S and T, the fact that, say, S is published means that the Speaker is in the dataset. This violates differential privacy: under differential privacy, if S can be published when the Speaker is present, then S should be published with similar probability (off by at most an e^{ϵ} factor), when the Speaker is absent.

provide specific details about other flavors of differential privacy, potentially leading to more exploration of differential privacy in general.

The Epsilon Registry exposes information that is essential to understanding the privacy afforded by differentially private systems. While we did not set out to tackle the issue of interpretability and explainability broadly, the components of the Epsilon Registry provide a concrete example of the sort of information that must be divulged for experts to reason about black box algorithmic systems. It so happens that a comprehensive evaluation of differential privacy in practice requires knowledge of many moving pieces – data characteristics, ϵ per datum, burn rate, etc. – all of which are unknown to individuals outside an organization.

The Registry contents compares favorably to the Outcome-Logic-Design framework proposed by Selbst and Barocas [2017]. The framework distinguishes between information about the factors that led to a specific decision (outcome); the rules and guidelines used in the process as whole (logic); and the design choices, assumptions made by the developers, etc. (design). The Registry captures information at each layer. In particular, at the outcome layer, the factors that contributed to a differentially private result are computationally related to the input data (factors of which are captured in the Registry) and the choice of ϵ. The flavor of differential privacy used discloses information about the logic layer. Finally, the disclosures about the methods and justifications for implementation choices provide information about the design layer.

While the technical content of the Registry may be incomprehensible to the general public, the Registry provides information that allows those familiar with differential privacy to evaluate the privacy protection that differentially private systems provide. As with any concept, understanding is dependent on the recipient's knowledge [Ananny and Crawford, 2018].[22] Finally, in the right hands, the contents of the Registry could also facilitate accountability. Using the Registry information, experts could call out firms making privacy claims involving differential privacy in the marketplace but offering implementations that deliver little to no privacy (e.g. when organization use high ϵs or utilize resetting budgets).

The Epsilon Registry speaks to the broader question of how to address system-level opacity, a hotly contested topic in the growing field of fair, transparent, explainable and accountable machine learning [Kluttz et al., 2018, Burrell, 2016]. It also raises a set of common high-level questions that must be answered to explore a system's impact on values. When considering issues of fairness, a common question that arises is "fair with respect to what or whom?" In this paper, we encountered the analog – "differentially private with respect to what or whom?"[23] When considering questions of interpretability, a question that arises is "interpretable to whom?" For our case, the goal of the Registry was to improve experts' understanding.[24] As such, this paper serves as a case study for what is required to scrutinize an opaque system.

[22] Shah and Kesan [2003] note this idea of "transparency to whom" - however, their use of transparency relates to both the disclosure and the understandability of information; nonetheless, the same idea still holds - understanding is dependent on the individual in question.

[23] The question of "differentially privacy with respect to what or whom" is better defined than the question of "fair with respect to what or whom."

[24] At first glance, this may seem limited. However, the interviews from our census provide a reason for optimism. We observed that experiential learning via simulations was fruitful in gaining buy-in within organizations by somewhat filling the knowledge gap between differential privacy practitioners and their collaborators. While we cannot say that this approach works in all situations, our interviews provide an existence case of expanding the interpretability of a technically sophisticated process to a wider audience.

In relation to the role of explanations, much of the calls to action consider the disclosure of the dynamics of a technical system in a vacuum. They focus solely on the regime governing the system's decision making, without considering the external factors that went into its construction. Our research and analysis suggest that scrutinizing a system solely on the technical specifications is not sufficient. Comprehensive analysis of the entire environment – inputs, scope, unit of analysis, the choices that human actors made throughout this process – need to be disclosed and understood as well. Selbst and Barocas' Outcome-Logic-Design framework speaks to this very point. The outcome and logic design emphasize explanation of the system level mechanics for individual outputs and in general, while the design layer acknowledges the contribution of non-system forces into a system's design. The scope of our findings can be applied more broadly to systems and algorithms where contextual nuances are ingrained throughout the development cycle.

7. Conclusion

On a fundamental level, a contribution to the Epsilon Registry is a system-level transparency report. It provides the necessary information required for individuals with knowledge of differential privacy to assess the quality of privacy afforded. As time progresses, and new techniques emerge, it may be useful to contribute new components to the Registry.

The information we identified as necessary to evaluate the privacy afforded by a differentially private system provides some insight into the types of information necessary to support evaluations of algorithmic systems more generally. Design choices and justifications, as well as the interactions between an algorithm and the data, were all required to evaluate the system. Further engagement with the literature and practices around reproducible research may advance research and practice around transparency. In particular, while we chose to focus on methodological justifications here, the other components of reproducible research that highlight data and code are important as well.

Absent additional incentives, it is unlikely that institutions will voluntarily disclose information to the Registry. We argued that an information disclosure law is well-suited to remedy this misalignment of incentives by opening up closed practices to greater scrutiny.

Differential privacy presents an opportunity to make objective, verifiable statements about privacy loss. This is a rare event in a field largely dominated by unsubstantiated and unverifiable claims. Adoption of the Epsilon Registry would spur more meaningful implementations of differential privacy, in the moment and over time. Better differentially private implementations will reduce the risks to privacy posed by statistical learning in the public and private sector. This in turn, will allow society to reap the benefits of big data while protecting individual and collective interests in privacy.

References

J. M. Abowd. The U.S. Census Bureau Adopts Differential Privacy. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2018.

A. L. Allen. Coercing Privacy. *Wm. & Mary L. Rev.*, 40:723, 1998.

M. Ananny and K. Crawford. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 20(3):973–989, 2018.

M. Bun and T. Steinke. Concentrated Differential Privacy: Simplifications, Extensions, and Lower Bounds. In *Theory of Cryptography Conference*, pages 635–658. Springer, 2016.

J. Burrell. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1):2053951715622512, 2016.

J. A. Calandrino, A. Kilzer, A. Narayanan, E. W. Felten, and V. Shmatikov. "You Might Also Like:" Privacy Risks of Collaborative Filtering. In *2011 IEEE Symposium on Security and Privacy*, pages 231–246. IEEE, 2011.

I. Dinur and K. Nissim. Revealing Information while Preserving Privacy. In *Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, pages 202–210. ACM, 2003.

C. Dwork and G. N. Rothblum. Concentrated Differential Privacy. *arXiv preprint arXiv:1603.01887*, 2016.

C. Dwork, A. Smith, T. Steinke, and J. Ullman. Exposed! A Survey of Attacks on Private Data. *Annual Review of Statistics and Its Application*, 4:61–84, 2017.

Ú. Erlingsson, V. Pihur, and A. Korolova. Rappor: Randomized Aggregatable Privacy-Preserving Ordinal Response. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, pages 1054–1067. ACM, 2014.

R. Gellman. Fair Information Practices: A Basic History. *Available at SSRN 2415020*, 2017.

R. M. Groves, M. E. Chernew, P. Daas, C. Dwork, O. Frieder, B. Harris-Kojetin, H. V. Jagadish, F. Kreuter, S. Lohr, J. P. Lynch, C. O'Muircheartaigh, T. Raghunathan, R. Rigobon, and M. Rotenberg. *Federal Statistics, Multiple Data Sources, and Privacy Protection: Next Steps*. National Academies Press, 2017.

N. Homer, S. Szelinger, M. Redman, D. Duggan, W. Tembe, J. Muehling, J. V. Pearson, D. A. Stephan, S. F. Nelson, and D. W. Craig. Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays. *PLoS genetics*, 4(8):e1000167, 2008.

D. Kluttz, N. Kohli, and D. K. Mulligan. Contestability and Professionals: From Explanations to Engagement with Algorithmic Systems. *Available at SSRN 3311894*, 2018.

A. Korolova. Privacy Violations Using Microtargeted Ads: A Case Study. In *Data Mining Workshops (ICDMW), 2010 IEEE International Conference on*, pages 474–482. IEEE, 2010.

R. Kumar, J. Novak, B. Pang, and A. Tomkins. On Anonymizing Query Logs via Token-Based Hashing. In *Proceedings of the 16th international conference on World Wide Web*, pages 629–638. ACM, 2007.

B.-C. Lee, L. Ang, and C. Dubelaar. Lemons on the Web: A signalling approach to the problem of trust in Internet commerce. *Journal of Economic Psychology*, 26(5):607–623, 2005.

A. Machanavajjhala, D. Kifer, J. Abowd, J. Gehrke, and L. Vilhuber. Privacy: Theory meets Practice on the Map. In *Proceedings of the 2008 IEEE 24th International Conference on Data Engineering*, pages 277–286. IEEE Computer Society, 2008.

W. A. Magat, W. A. Magat, W. K. Viscusi, and W. A. Magat. *Informational Approaches to Regulation*, volume 19. MIT press, 1992.

I. Mironov. Renyi Differential Privacy. In *Computer Security Foundations Symposium (CSF), 2017 IEEE 30th*, pages 263–275. IEEE, 2017.

A. Narayanan and V. Shmatikov. Robust De-anonymization of Large Sparse Datasets. In *Security and Privacy, 2008. SP 2008. IEEE Symposium on*, pages 111–125. IEEE, 2008.

I. Rahwan. Society-in-the-Loop: Programming the Algorithmic Social Contract. *Ethics and Information Technology*, 20(1):5–14, 2018.

A. D. Selbst and S. Barocas. Regulating Inscrutable Systems. In *WeRobot 2017*, 2017.

R. C. Shah and J. P. Kesan. Manipulating the Governance Characteristics of Code. *Info*, 5 (4):3–9, 2003.

V. Stodden. Reproducible research for scientific computing: Tools and strategies for changing the culture. *Computing in Science & Engineering*, 14(4):13–17, 2012.

V. Stodden and S. Miguez. Best practices for computational science: Software infrastructure and environments for reproducible and extensible research. *Available at SSRN 2322276*, 2013.

C. R. Sunstein. Informing America: Risk, Disclosure, and the First Amendment. *Fla. St. UL Rev.*, 20:653, 1992.

C. R. Sunstein. Informational Regulation and Informational Standing: Akins and Beyond. *U. Pa. L. Rev.*, 147:613, 1998.

L. Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570, 2002.

J. Turow, C. J. Hoofnagle, D. K. Mulligan, and N. Good. The Federal Trade Commission and Consumer Privacy in the Coming Decade. *ISJLP*, 3:723, 2007.

A. Wood, M. Altman, A. Bembenek, M. Bun, M. Gaboardi, J. Honaker, K. Nissim, D. R. O'Brien, T. Steinke, and S. Vadhan. Differential Privacy: A Primer for a Non-Technical Audience. *Vand. J. Ent. & Tech. L.*, 21:209, 2018. URL http://www.jetlaw.org/journal-archives/volume-21/volume-21-issue-1/differential-privacy-a-primer-for-a-non-technical-audience/.

Appendix A. Canon of Counterexamples

The following is a list of counterexamples to presumed confidentiality protection:

- Sweeney [2002]'s re-identification of Governor William Weld's medical record information showed that the combination of several attributes (in this case, race, birthdate, and zip code) can uniquely identify, or "name", an individual in the dataset.
- Narayanan and Shmatikov [2008] made the observation that a small collection of movie titles and the approximate dates on which they were watched were sufficient to uniquely identify most of the 480,000 individuals in the Netflix prize training set.
- Calandrino et al. [2011] describe how to exploit the relative timing of a consumer's blog posts and changes in the outputs of a product recommendation system to infer purchases made, but not blogged about, by the blogger.
- Korolova [2010] created several Facebook advertising campaigns that paired sufficiently many known attributes of the same specific individual to name her with various age ranges. The campaign for which Korlova was charged revealed the individual's age range.
- Kumar et al. [2007] combined a statistical analysis of "tokenized" search logs, in which each term t had been replaced by a hash $h(t)$ with an older and disjoint public search log (specifically, the published log from the famous AOL search debacle) to effectively reverse engineer the hash function and reconstruct the searches, showing that tokenization does not preserve privacy of searches.
- Homer et al. [2008] gave a procedure for testing an individual's membership in a Genome-Wide Association Study (GWAS) case group given (1) aggregate statistics for allele frequencies in the case group; (2) aggregate statistics for allele frequencies for an ethnically

similar healthy population, available, for example, from HapMap (or which could be available from a control group in the same study), and the DNA of the individual. This resulted in aggregate allele frequencies in NIH-funded research being withheld from the general public. See Homer et al. [2008] and the survey of Dwork et al. [2017].

Appendix B. Interview Questionnaire

(1) *Why Differential Privacy*:
   - What was the impetus?
   - Was there a particular adversary that motivated the adoption?
   - Was there a particular event?
   - Where there any barriers?

(2) *Data Characteristics & Avenues for Loss*:
   - Describe the use of the data: what are the avenues for privacy loss?
   - What are the uses of the data during which information regarding one person can affect the experience or be exposed to another person?

(3) *Granularity of Protection*: At what level of data granularity are you providing/measuring protection?

(4) *Limitations on Privacy Loss*:
   - Do you have any limits on privacy loss?
   - If so, over what time period or collection of exposures do the limits apply?

(5) *Algorithm Specifics*: These questions were interested in the parameter specifications of the differentially private algorithm. For example:
   - What variant of differential privacy is in use? Why was it selected? What was the process of selection? Who was involved? What factored into consideration?
   - What is your burn rate: What value of ϵ (and other parameters) per lifetime? Time period? Why was it selected? What was the process of selection? Who was involved? What factored into consideration?
   - If you don't guarantee a lifetime bound on the privacy loss, what is your mitigation for repeated use? [This is something into which we will want to drill down – what is their intuition regarding the mitigation.]
   - Are data "retired" after they have reached a certain degree of exposure, e.g., worst-case or expected cumulative privacy loss? What is the threshold for retirement?