NBER WORKING PAPER SERIES

PRIVACY AND DATA-BASED RESEARCH

Ori Heffetz
Katrina Ligett

Working Paper 19433
http://www.nber.org/papers/w19433

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2013

For useful comments on an early draft, we thank Dan Benjamin, Avrim Blum, Hank Greely, Aleksandra Korolova, Frank McSherry, Kobbi Nissim, Ted O'Donoghue, Grant Schoenebeck, Moses Shayo, Latanya Sweeney, Kunal Talwar, and Jonathan Ullman. Ligett's work was supported in part by an NSF CAREER award (CNS-1254169), the US-Israel Binational Science Foundation (grant 2012348), the Charles Lee Powell Foundation, a Google Faculty Research Award, and a Microsoft Faculty Fellowship. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Privacy and Data-Based Research
Ori Heffetz and Katrina Ligett
NBER Working Paper No. 19433
September 2013
JEL No. C49,C89,D89,Z00

ABSTRACT

What can we, as users of microdata, formally guarantee to the individuals (or firms) in our dataset, regarding their privacy? We retell a few stories, well-known in data-privacy circles, of failed anonymization attempts in publicly released datasets. We then provide a mostly informal introduction to several ideas from the literature on differential privacy, an active literature in computer science that studies formal approaches to preserving the privacy of individuals in statistical databases. We apply some of its insights to situations routinely faced by applied economists, emphasizing big-data contexts.

Ori Heffetz
S.C. Johnson Graduate School of Management
Cornell University
324 Sage Hall
Ithaca, NY 14853
and NBER
oh33@cornell.edu

Katrina Ligett
Department of Computing and Mathematical Sciences
and
Division of Humanities and Social Sciences
California Institute of Technology
MC 305-16
Pasadena, CA 91125
katrina@caltech.edu

On August 9, 2006, the Technology section of the *New York Times* contained a news item titled "A Face Is Exposed for AOL Searcher No. 4417749." In it, reporters Michael Barbaro and Tom Zeller tell a story about big data and privacy:

> Buried in a list of 20 million Web search queries collected by AOL and recently released on the Internet is user No. 4417749. The number was assigned by the company to protect the searcher's anonymity, but it was not much of a shield.
>
> No. 4417749 conducted hundreds of searches over a three-month period on topics ranging from "numb fingers" to "60 single men" to "dog that urinates on everything."
>
> And search by search, click by click, the identity of AOL user No. 4417749 became easier to discern. There are queries for "landscapers in Lilburn, Ga," several people with the last name Arnold and "homes sold in shadow lake subdivision gwinnett county georgia."
>
> It did not take much investigating to follow that data trail to Thelma Arnold, a 62-year-old widow who lives in Lilburn, Ga.
>
> . . .
>
> Ms. Arnold, who agreed to discuss her searches with a reporter, said she was shocked to hear that AOL had saved and published three months' worth of them. "My goodness, it's my whole personal life," she said. "I had no idea somebody was looking over my shoulder."
>
> . . .
>
> "We all have a right to privacy," she said. "Nobody should have found this all out."

Empirical economists are increasingly users, and even producers, of large datasets with potentially sensitive information. Some have for decades handled such data (e.g., Census data), and routinely think and write about privacy. Many others, however, are not accustomed to think about privacy, perhaps because their research traditionally relies on already-publicly-available data, or because they gather their data through relatively small, "mostly harmless" surveys and experiments. This ignorant bliss may not last long; detailed data of unprecedented quantity and accessibility are now ubiquitous: a private database from an internet company, field experimental data on massive groups of unsuspecting subjects, or a government agency's confidential administrative records in digital form. And while big data become difficult to avoid, getting privacy right is far from easy—even for data scientists, as the AOL story demonstrates.

This paper aims to inspire data-based researchers to think more about issues such as privacy and anonymity. Many of us routinely promise anonymity to the subjects who participate in our studies, either directly through informed consent procedures, or indirectly through our correspondence with Institutional Review Boards (IRBs). What is the informational content of such promises? Given that our goal is, ultimately, to publish the results of our research—formally, to publish functions of the data—under what circumstances, and to what extent, can we guarantee that individuals' privacy not be breached and anonymity not be compromised?

These questions may be particularly relevant in a big data context, where there may be a risk of more harm—due to often-sensitive content—to more people—due to scale. As we discuss below, it is also in a big data context that "privacy guarantees" of the sort we consider may be most effective.

Our paper is divided into three parts. In the first, we retell the stories of several privacy debacles, well-known in privacy-research circles. The first three stories concern intentional releases of deidentified data for research purposes. In each case attempts were made to anonymize the data—records of medical patients, web searchers (the AOL story above), and Netflix subscribers—prior to their release, but it quickly became apparent that individuals could be reidentified. The fourth story—involving Facebook ads—illustrates how individuals' privacy could be breached even when the data themselves are not released. It demonstrates that a potential privacy risk exists whenever some *function* of personal data—in this case, the seemingly innocuous count of ads shown in a campaign—is visible to outsiders—in this case, advertisers. These often-told stories serve as cautionary tales of how things can go terribly wrong when, in the absence of a formal framework for thinking about privacy, one relies instead on intuition. We also discuss a number of natural, more sophisticated proposals and attempts to disguise the data in each of the four cases; while making reidentification more difficult, they do not eliminate the risk.

None of our stories involves *security* horrors such as stolen data, broken locks and passwords, or compromised secure connections. Rather, in all of them information was released that had been *thought* to have been anonymized, but, as was quickly pointed out, was rather revealing. A naive reaction might be to attempt to roll back current policies and reverse current trends of increased openness and data sharing. We believe such a reaction would be misguided. Moreover, as the Facebook example demonstrates, not sharing the data does not eliminate the risks.

In the second part of our paper we shift gears, switching from storytelling to discussing *differential privacy*, a rigorous, portable privacy notion introduced roughly a decade ago by computer scientists aiming to enable the release of information while providing *provable* privacy guarantees. We formally introduce the differential privacy definition, at the heart of which is the idea that the addition or removal of a single individual from a dataset should have nearly no effect on any publicly released functions of the data—a single statistic, a collection of statistics, synthetic data that preserve many properties of the original database, or a complex policy recommendation. Achieving this goal requires introducing randomness into the released outcome, with more randomness resulting in more privacy—but less accuracy. We discuss simple applications of the definition, illustrating this tension.

What insights can the differential privacy literature offer regarding our cautionary tales? What guidance does it have for researchers working with data? In the third part of our paper, we offer lessons and reflections, discuss some limitations, and briefly mention additional applications. We conclude with reflections on current promises of "anonymity" to study participants—promises that, given common practices in empirical research, are not guaranteed to be kept. We invite researchers to consider either backing such promises with meaningful privacy-preserving techniques, or qualifying them. Especially in the case of big data, application-ready implementations of such techniques may in the foreseeable future become a practical possibility, but only if increased awareness among data-based researchers generates demand. Until then, more cautious promises may be warranted; while we are not aware of major privacy debacles in economics research to date, the stakes are only getting higher.

Intuition Regarding Privacy May Not Be Enough

Well-intentioned government or private entities in possession of a sensitive database may wish to make an anonymized version of it public, for example to facilitate research. We retell and discuss a few cautionary tales that illustrate how intuition-based anonymization attempts may fail, sometimes spectacularly.[1]

[1] As mentioned above, these stories are "classics," well known in the computer science community that studies privacy, and routinely cited in the introduction to many papers. The first three were recently revisited and discussed by Ohm (2010), a legal scholar, who provides further references and links to primary sources.

Stories of Failed Anonymization

The first story is from the mid 1990s, when William Weld, then Governor of Massachusetts, approved the release of certain medical records of state employees to researchers, assuring the public that individual anonymity would be protected by eliminating obvious identifiers from the data (Greely, 2007). A few days after Weld's announcement, Latanya Sweeney—then a graduate student at MIT—reidentified Weld's personal records (including diagnoses and prescriptions) in the database; she then had his records delivered to his office.

While the medical data—officially, the Massachusetts "Group Insurance Commission" (GIC) data—had been "deidentified" by removing fields containing patients' name, address, and social security number (SSN) prior to the the data release, the nearly one hundred remaining fields included ZIP code, birth date, and sex. As Ohm (2010) tells the story, Sweeney

> knew that Governor Weld resided in Cambridge, Massachusetts, a city of fifty-four thousand residents and seven ZIP codes. For twenty dollars, she purchased the complete voter rolls from the city of Cambridge—a database containing, among other things, the name, address, ZIP code, birth date, and sex of every voter. By combining this data with the GIC records, Sweeney found Governor Weld with ease. Only six people in Cambridge shared his birth date; only three were men, and of the three, only he lived in his ZIP code.[2,3]

The next story, involving Ms. Arnold above, is from roughly a decade later. In 2006, AOL Research released detailed internet search records of 650,000 users covering a three-month period, amounting to twenty million search queries.[4] The stated purpose of the release was expressed by

[2] Barth-Jones (2012) revisits and critiques this story. Perhaps in response, Sweeney, Abu and Winn (2013) use a similar method to reidentify individuals in the publicly available Personal Genome Project database.

[3] Sweeney's "How Unique Are You?" interactive website invites the visitor to "Enter your ZIP code, date of birth, and gender to see how unique you are (and therefore how easy it is to identify you from these values)." Her simple methodology is explained on the website: (`http://aboutmyinfo.net`, accessed on August 9, 2013)

> "Birthdate . . . , gender, and 5-digit postal code (ZIP) uniquely identifies most people in the United States. Surprised? . . . 365 days in a year x 100 years x 2 genders = 73,000 unique combinations, and because most postal codes have fewer people, the surprise fades. . . . there are more than 32,000 5-digit ZIP codes in the United States; so 73,000 x 32,000 is more than 2 billion possible combinations but there are only 310 million people in the United States. In 1997, Latanya Sweeney did this kind of uniform calculation on populations reported in the U.S. Census for age groups in each postal code and summed the results to predict that at most 87 percent of the U.S. population had unique combinations . . . the maximum percent of unique combinations may drop as you move from date of birth to age, and from 5-digit ZIP code to county. Notice that even knowing the county, age, and gender can make some people unique. They are few, and they tend to live in remote locations, but notice it is not 0."

Golle (2006) revisits and updates this 87 percent figure.

[4] As Ohm (2010) notes, different numbers appear in different accounts. The 650,000 figure above was described

then AOL Research head Abdur Chowdhury:

> AOL is embarking on a new direction for its business - making its content and products freely available to all consumers. To support those goals, AOL is also embracing the vision of an open research community, which is creating opportunities for researchers in academia and industry alike. ... with the goal of facilitating closer collaboration between AOL and anyone with a desire to work on interesting problems.[5]

Prior to the data release, the search logs were deidentified, for example by removing usernames and IP addresses, using instead unique identifiers (such as "4417749") to link all of a single user's queries. This deidentification, however, quickly proved far from sufficient for the intended anonymization, as illustrated by the *New York Times* article on Ms. Arnold. Within days of the release, AOL apologized, removed the data website as well as a few employees, and silenced its Research division. Of course, to this day, the data are widely available through a simple web search; once published, you cannot take it back.

The third story is also from 2006, a bad year for privacy. About two months after the AOL debacle, Netflix announced a competition—the Netflix Prize—for improving the company's algorithm that predicts user ratings of films, using only past user ratings. To allow competitors to train their algorithms, Netflix released a database with one hundred million ratings of 17,770 films by half a million subscribers covering a six-year period. Each record contained a movie title, a rating date, and a five-point rating. As in the GIC and AOL cases, records were deidentified prior to the release, replacing user names with unique identifiers.

The illusion of protecting users' anonymity was, again, short-lived. Two weeks after the data release, Narayanan and Shmatikov (2008; first version posted in 2006) demonstrated that, in their (2008) words, "an adversary who knows a little bit about some subscriber can easily identify her record if it is present in the dataset, or, at the very least, identify a small set of records which include the subscriber's record."

How little is "a little bit"? In many cases, as little as knowing a user's approximate dates and ratings of two or three movies. In their demonstration, Narayanan and Shmatikov used ratings from the Internet Movie Database (IMDB), which are publicly available and are linked to the raters'

as 500,000 in the original post, and the twenty million figure in the original post has later been reported by some as thirty-six million.

[5] Posting of Abdur Chowdhury, cabdur@aol.com, to SIGIR-IRList, `irlist-editor@acm.org`, `http://sifaka.cs.uiuc.edu/xshen/aol/20060803_SIG-IRListEmail.txt` (cited in Ohm (2010), and accessed on August 9, 2013).

identities, and showed how a handful of a user's IMDB ratings, even when they yield imprecise information, could uniquely identify her in the Netflix database.

Whereas IMDB's public ratings may reveal only those movies that an individual wants the world to know she has watched, Netflix ratings may reveal *all* of the movies she has rated, including those she may prefer to keep private—e.g., films that may reflect her sexual, social, political, or religious preferences. Moreover, to be reidentified, one does not have to be on IMDB: as Ohm (2010) advises his readers, "the next time your dinner party host asks you to list your six favorite obscure movies, unless you want everybody at the table to know every movie you have ever rated on Netflix, say nothing at all."

Deidentification and Beyond

Deidentified data were defined by Sweeney (1997) as data in which "all explicit identifiers, such as SSN, name, address, and telephone number, are removed, generalized, or replaced with a made-up alternative." Her definition seems to accurately describe the released GIC, AOL, and Netflix data in the stories above. Some more recent definitions (e.g., those under federal health records privacy regulations) are stricter and would not consider the GIC data released by Weld deidentified, but they too keep the focus on removing only specific kinds of information (Greely, 2007). Indeed, more than fifteen years after Sweeney's powerful demonstration, her definition *still* describes, more or less accurately, commonplace practices among many researchers. For example, prior to publicly posting their data online (as required by some journals), economists often deidentify their data by merely withholding explicit identifiers such as subject names. However, as in the stories above, the stated aim of such deidentification—and what is often promised to subjects, directly or via IRB—is *anonymization.* In Sweeney's (1997) definition, *anonymous data* "cannot be manipulated or linked to identify an individual." Clearly, deidentification far from guarantees anonymization.

Sweeney's GIC reidentification used birthday and 5-digit ZIP code, neither of which are typically included in datasets publicly posted by economists. But it is not difficult to imagine reidentification of specific subjects based on combinations of demographics such as study major, age/class, gender, and race, which are often *not* considered "identifiable private information" and are routinely included in posted data.[6] Reidentification is still easier with knowledge regarding, e.g., the

[6] For example, according to Cornell's Office of Research Integrity and Assurance:

> "Identifiable private information is defined as: name; address; elements of dates related to an individual (e.g., birth date); email address; numbers: telephone, fax, social security, medical record, health beneficiary / health insurance, certificate or license numbers, vehicle, account numbers (e.g., credit

day and time in which a classmate or a roommate participated in a specific study session.[7] But it is possible even without such special knowledge, and it may be straightforward for specifically vulnerable individuals, such as minorities, or women in the sciences.

This discussion highlights a weakness of deidentification: if one assumes no restrictions on outside information (also referred to below as auxiliary information), then, short of removing *all* data fields prior to a release, some individuals may be uniquely identified by the remaining fields. One potential response to this weakness is an approach called *k-anonymity*, which combines *some* restrictions on outside information with the removal (or partial removal) of *some* fields. Specifically, assuming that outside information could only cover certain fields in the database, one could suppress these fields or, when possible, generalize them (e.g., replace date of birth with year of birth), so that any combination of the values reported in these fields would correspond to at least k individuals in the data (see, e.g., Sweeney, 2002). This approach has several weaknesses, and in many applications it implies either an unreasonably weak privacy guarantee or a massive suppression of data (notice that the amount of information that can be released is expected to shrink as k grows and as restrictions on outside information are weakened). See, e.g., Narayanan and Shmatikov (2008) for a discussion in the Netflix context.

An alternative approach is to make it harder for an attacker to leverage outside information. For example, prior to making the query logs publicly available, AOL could have replaced not only user identities but also *the search words themselves* with uniquely identifying random strings. Similarly, Netflix could have replaced movie names with unique identifiers. Such an approach, known as "token-based hashing," would preserve many features of the data, hence maintaining usefulness of the database for some (though, clearly, not all) research purposes. But it is these very preserved features of the underlying data that make this scheme vulnerable as well.

Indeed, shortly after the disaster at AOL Research, a group at Yahoo! Research—Kumar et al. (2007)—showed that an attacker with access to a "reference" query log (e.g., early logs released by Excite or Altavista) could use it to extract statistical properties of tokenized words in the database, and "invert the hash function," i.e., break the coding scheme, based on co-occurrences of tokens within searches. Along similar lines, Narayanan and Shmatikov (2008) speculate that in the Netflix

> card), device identification numbers, serial numbers, any unique identifying numbers, characteristics, or codes (e.g., Global Positioning System (GPS) readings); Web URLs; Internet Protocol (IP) addresses; biometric identifiers (e.g., voice, fingerprints); full face photographs or comparable images."

(`http://www.irb.cornell.edu`, accessed on August 13, 2013).

[7] In a similar vein, Sweeney (2013) uses newspaper stories that contain the word "hospitalized" to reidentify individual patients in a publicly available health dataset in Washington State.

case, such an approach “does not appear to make de-anonymization impossible, but merely harder.”

Privacy Risk Without Data Release

Our fourth story, of privacy compromised on Facebook by Korolova (2011),

> illustrates how a real-world system designed with an intention to protect privacy but without rigorous privacy guarantees can leak private information … Furthermore, it shows that user privacy may be breached not only as a result of data publishing using improper anonymization techniques, but also as a result of internal data-mining of that data.

Facebook’s advertising system allows advertisers to specify characteristics of individuals to whom an ad should be shown. At the time of Korolova’s attack, it was possible to specify those characteristics (e.g., gender, age, location, workplace, alma mater) so finely that they would correspond to a unique Facebook user. Then, two versions of the ad campaign could be run—for example, one with those same characteristics plus “Interested in women;” the other with those characteristics plus “Interested in men.” Even if this user’s interests were not visible to her friends, if she had entered them in her profile, they would be used for ad targeting. Thus, if the advertiser received a report that, e.g., the “Interested in women” version of her ad had been displayed, she could infer the targeted individual’s private interests.

Other attacks were possible too. “Using the microtargeting capability, one can estimate the frequency of a particular person’s Facebook usage, determine whether they have logged in to the site on a particular day, or infer the times of day during which a user tends to browse Facebook.” Korolova’s paper quotes failed promises by Facebook executives, such as that Facebook “[doesn’t] share your personal information with services you don’t want” and “[doesn’t] give advertisers access to your personal information.” She notes:

> We communicated our findings to Facebook on July 13, 2010, and received a very prompt response. On July 20, 2010, Facebook launched a change to their advertising system that made the kind of attacks we describe much more difficult to implement in practice, even though, as we discuss, they remain possible in principle.

This Facebook story helps demonstrate that if one seeks to use functions of the data—be it via research findings, policy decisions, or commercial services and products—the privacy of the individuals comprising the data may be at risk without an approach providing (provable) privacy guarantees. We discuss such an approach next.

Differential Privacy

A common theme in the examples above has been the crucial role played by *auxiliary information*, i.e., knowledge from sources *outside* the dataset under consideration. In the examples above, attackers consulted various outside sources not foreseen by the database owners—public records such as voter rolls, complementary databases such as IMDB, or, simply, personal familiarity with an individual in the database. To identify individuals, the attackers then carried out a variant of a so-called "linkage attack:" they matched fields that overlap across the auxiliary data and the attacked database.

More generally, one may invite trouble when making specific assumptions regarding *what* information a potential attacker might have or *how* she might use it. If such assumptions are ever violated—even in the future, as new technology and information become available—privacy may be compromised. One approach to addressing the auxiliary-information concern would be to seek to provide privacy guarantees free from assumptions about the attacker and her information.

The approach we discuss here, *differential privacy*, seeks to do just that. It emerged from work in computer science theory by Dinur and Nissim (2003), Dwork and Nissim (2004), and Dwork et al. (2006*a*). Our discussion and examples draw on a number of surveys, including Dwork (2006), Dwork and Smith (2010), Dwork (2011*b*,*a*), and Dwork et al. (2011). These surveys additionally present historical aspects of the development of the differential privacy definition, more examples, and a much broader range of applications than we discuss here.[8]

The Differential Privacy Definition

To fix ideas, consider the released outcome of some function of a database (for example, the released number of Facebook users to whom an ad was displayed, or some published table of statistics in an empirical research paper, or even a released version of the entire database). Consider a potential participant in the database (for example, someone who considers joining Facebook, or someone who considers participating in a research study), and compare two possible scenarios: in one, she joins and is added to the database; in the other, she does not join and is hence not in the database.

[8] We also recommend a recent popular article on differential privacy research by Klarreich (2012).

Informally, differential privacy seeks to guarantee to the potential participant that, irrespective of her participation decision, *almost* the same things can be learned from the released outcome—regardless of outside information, of the data already in the database, or of her own personal data. We emphasize "*almost*"; without it, the function would have to yield identical results on any two input databases—even two that are wildly different—since any database of individuals can be constructed from any other via a (potentially long) chain of additions and removals of individuals.

Differential privacy hence gives participants (and nonparticipants) in the database a form of plausible deniability: they could always deny that their data took specific values or even that they participated (or not), and an observer would have almost no evidence either way.

Specifically, consider pairs of databases (D, D') that are identical except that one of the databases has one additional row (or record) over the other. We refer to such a pair as *neighboring* databases, and think of each row as corresponding to one individual. Thus, two neighboring databases differ by only the participation of one individual. Now consider some computation on such databases. Denote the computation by a function K, and consider the space of possible outcomes of the computation. A differentially private computation (or function, or mechanism) selects its output using randomness, such that the probability of any given outcome is similar under any two neighboring databases.

Formally,

Definition (**Differential Privacy**; Dwork et al., 2006*a*)**.** *A randomized function K provides ϵ-differential privacy if for every $S \in \mathit{Range}(K)$ and for all neighboring databases D and D',*

$$\text{Prob}[K(D) = S] \leq e^{\epsilon} \cdot \text{Prob}[K(D') = S],$$

for $\epsilon \geq 0$, and where the probability space in each case is over the randomness of K.[9]

Note that in particular, for any neighboring pair (D, D'), the definition must hold with the larger quantity (i.e., $\max\{\text{Prob}[K(D) = S], \text{Prob}[K(D') = S]\}$) on the left, constraining it to be larger by at most a multiplicative e^{ϵ}.

This definition formalizes our discussion above: by "*almost* the same" we meant that ϵ should be small. How small? The definition does not prescribe an answer to this normative question,

[9] There are other variants on this definition, which we do not emphasize here. A common generalization of differential privacy allows an *additive* δ difference in the probabilities, in addition to the mulitiplicative difference e^{ϵ} (see, e.g., Dwork et al., 2006*b*; Machanavajjhala et al., 2008). Such generalization provides a weaker privacy guarantee, but may allow for more accurate outcomes.

a point we return to below. In the limiting case of $\epsilon = 0$, we would drop the "*almost*"; but in that limiting case, e^ϵ would equal 1, requiring that the function K be indistinguishable on any two input databases, as discussed above. In other words, maximum differential privacy means useless published output. More generally, the definition makes precise an intuitive tradeoff between privacy and usefulness.

The definition readily extends to provide a privacy guarantee to a group of individuals: an ϵ-differentially private mechanism is $k\epsilon$-differentially private from the point of view of a group of k individuals (or one individual whose data comprise k rows in the database). It also immediately yields an elegant composition property: running l ϵ-differentially private mechanisms—e.g., publishing l statistics based on the database—gives a guarantee of $l\epsilon$-differential privacy.[10] This composition property is particularly important in the context of academic research, where individuals may participate in more than one database, and where on each database typically more than one analysis is conducted. Differential privacy hence provides a tool for understanding the cumulative privacy harm incurred by an individual whose data appear in multiple databases, potentially used by different entities for different purposes and at different points in time. One could discuss assessments of individuals' cumulative, lifelong privacy loss, and use them as an input into the discussion of how small ϵ should be.

Finally, it is also useful to note that differential privacy guarantees hold up under postprocessing of their outputs: if one conducts an ϵ-differentially private computation, one is then free to perform any subsequent computation on its output, and the result will still be ϵ-differentially private. This means, for example, that once one has produced differentially private statistics on a dataset, those statistics can made public for all eternity, without concern that at some later date a clever hacker will find some new privacy-revealing weakness.

From a Bayesian point of view, differential privacy can be given the following interpretation: an observer with access to the output of a differentially private function should draw almost the same conclusions whether or not one individual's data are included in the analyzed database, regardless of the observer's prior. This interpretation highlights that differential privacy is a property of the function (the randomized mapping from databases into outcomes), not of the output (a particular outcome). Kasiviswanathan and Smith (2008) credit Dwork and McSherry with the first formulation of this interpretation, which can be formalized and proven equivalent to differential

[10] More generally, running any l differentially private mechanisms with guarantees $\epsilon_1, \ldots, \epsilon_l$ gives $\left(\sum_{i=1}^{l} \epsilon_i\right)$-differential privacy. Equivalently, one may split a fixed total budget of ϵ across a set of desired computations.

privacy.

"Observer" may of course refer to anyone with access to the output of the function, including, e.g., malicious attackers, (legitimate) advertisers on Facebook, or the readers of a research paper that reports some statistic. Notice that this Bayesian interpretation does *not* rule out performing analyses and reporting outcomes that vastly alter the adversary's posterior view of the world, so long as the outcomes are not very sensitive to the presence or absence of any one individual in the original database. An often-used example: one could conduct a differentially private analysis that revealed a surprising correlation between smoking and cancer, so long as that correlation depended only negligibly on the participation of any one individual (or, for that matter, of any small group of individuals) in the study. Revealing this correlation might allow observers to draw inferences about a smoker, who might then feel that her privacy has been harmed. But since essentially the same conclusions would have been drawn regardless of whether that smoker participated in the study, her *differential* privacy has been respected.

From Definition to Application: Noise and Sensitivity

Armed with the above definition, consider the computation (and subsequent release) of the mean income of individuals in a database. While the mean might seem like a fairly innocuous statistic, all statistics reveal *something* about the data, and there are worst-case situations where the mean might be quite revealing. For example, if the mean salary in a certain economics department prior to hiring a new faculty member is known to an observer (for instance, due to a previous release), then releasing the new mean after the hire reveals the new hire's salary. This is a variant of the so-called "differencing attack."

A differential privacy guarantee would require adding randomly generated noise to the true mean prior to its release. How much noise? Since differential privacy is a worst-case guarantee over all possible pairs of neighboring databases and over all possible outcomes, if the distribution of incomes is not a priori bounded, the noise would have to be unboundedly large (to guarantee that even the addition of an extreme outlier to the database would have little effect on the differentially-private statistic). With a limit on the range of incomes, however, one could add a limited amount of noise to the true mean in order to guarantee differential privacy.

Formally, when a function f that we wish to compute on a database returns a real number, we say that the *sensitivity* of that function is

$$\Delta f = \max_{D,D'} |f(D) - f(D')|,$$

for (D, D') neighboring databases. The definition makes it clear that sensitivity is a property of the function, given a universe of possible databases, and is independent of the actual input database. Intuitively, the sensitivity of f is simply the maximum difference between the values it takes on any two neighboring databases; this maximum difference must be hidden in order to preserve differential privacy.

A simple technique for hiding this maximum difference, Δf, in an ϵ-differentially private manner, is to add to f noise from a Laplace distribution with mean $= 0$ and standard deviation $= \sqrt{2}\Delta f/\epsilon$ (Dwork et al., 2006*a*).[11] We next focus on this technique in the context of a concrete example.

A Single-Statistic Example: Mean Salary

To illustrate some of the delicate issues involved in actually carrying out a differentially private computation, consider the release of mean salary among the faculty in an economics department. The technique above suggests calculating the (true) mean salary first, then adding to it Laplace noise with standard deviation $\sqrt{2}\Delta f/\epsilon$ prior to releasing it. One therefore needs, first, a value for ϵ. Such value could be determined through a combination of, e.g., philosophical and ethical inquiry, and social, political, and legislative processes, and could depend on context; further research is clearly needed.[12] Second, one needs to calculate Δf. In our case, this requires making assumptions regarding the universe of possible databases. We discuss these next.

Outcome range: As mentioned above, to yield practical results our technique requires Δf to be bounded. Our example intentionally involves salary, rather than total income, because salary

[11] With scale parameter $b = \Delta f/\epsilon$, the pdf of this distribution is $\frac{1}{2b}e^{-\frac{|x|}{b}}$ and its standard deviation is $\sqrt{2}b$. This distribution is a natural choice because its exponential form satisfies the multiplicative e^ϵ constraint in the differential privacy definition.

[12] As Dwork et al. (2011) note in a defense of differential privacy:

> Yes, this research is incomplete. Yes, theorems of the following form seem frighteningly restrictive:
>
> > If an individual participates in 10,000 adversarially chosen databases, and if we wish to ensure that her cumulative privacy loss will, with probability at least $1 - e^{-32}$, be bounded by e^1, then it is sufficient that each of these databases will be $\epsilon = 1/801$-differentially private.
>
> But how else can we find a starting point for understanding how to relax our worst-case adversary protection? How else can we measure the effect of doing so? And what other technology permits one to prove such a claim?

is bounded from below (in the worst case, at zero). One still needs an upper bound, which cannot be naively calculated from the data, but should be a property of the known universe of possible salaries. For simplicity, we assume that it is known to be some $\bar{y}$. With these bounds, and with mean salary as our outcome of interest, $|f(D) - f(D')| \leq \bar{y}/n$, where n is the number of individuals in the larger of the two neighboring databases.[13]

Database size: If the database size, n, is not publicly known, then the universe of possible databases includes the case $n = 1$, and therefore $\Delta f = \bar{y}$. With such high sensitivity, a naive application of the Laplace noise technique yields a uselessly uninformative outcome at any n: the noise added to the true mean has standard deviation $\sqrt{2}\bar{y}/\epsilon$, which, even with $\epsilon = 1$, is larger than the upper bound. An easy modification of the technique, however, yields noise that shrinks with n. The idea is to think of the mean as the function sum/n, treating n itself, as well as the sum of salaries, as two statistics to be calculated in a differentially private manner. One then divides the privacy budget ϵ between the two statistics: $\epsilon_n + \epsilon_{sum} = \epsilon$ (recall the composition property). Since the sensitivities of n and of the sum are, respectively, 1 and $\bar{y}$, the noise added would have standard deviations $\sqrt{2}/\epsilon_n$ and $\sqrt{2}\bar{y}/\epsilon_{sum}$. Since the two statistics increase with n, the noise-to-signal ratio of each vanishes asymptotically. With $\epsilon = 1$ and a favorable setting—a very large department with mean salary not much below $\bar{y}$—the differentially private release may convey some usable information about the true mean, but generally, the promise of the approach is more apparent on bigger data. For example, consider mean salary among the American Economic Association (AEA) membership ($n = 18,061$ in 2012).[14] With, e.g., $\epsilon = 0.1$ and true mean that is $\bar{y}/10$, the standard deviation on the noise added by the Laplace technique would be a much more tolerable 0.16% of the true n (i.e., 28 members) and 1.6% of the true sum of salaries, assuming we divide the privacy budget equally—rather than *optimally*—between the two statistics. Of course, things look still better with still bigger data and cleverer techniques.

Dwork (2011*a*) suggests that "[s]ometimes, for example, in the census, an individual's participation is known, so hiding presence or absence makes no sense; instead we wish to hide the values in an individual's row." Our examples above—of databases that include all faculty in an economics department or all AEA members—could be viewed and analyzed as settings where participation is publicly known. In such settings, it may make sense to modify our above definition of neighboring databases, from pairs "that are identical except that one of the databases has one additional row,"

[13] For simplicity (and conservativeness), in a database with $n = 0$ individuals we define mean salary to be at the lower bound 0.

[14] AEA membership figure is taken from Rousseau (2013).

to pairs of known size n, that differ in the content of exactly one row. In this form, differential privacy guarantees a participant that if her true salary y were replaced with some fake salary $y' \in [0, \bar{y}]$, the probability of any given outcome would not change by much. With this modification, only the sum of salaries needs to be computed and released in a differentially private manner.

Historically, this alternate definition (with databases of fixed and publicly known n) was used in the first papers that sparked the differential privacy literature, and it is still used in much of the work on differential privacy and statistics—a body of work that has grown quickly over the past few years. Work in this area has repeatedly established the feasibility of achieving common statistical goals while maintaining differential privacy. Importantly for our purposes, differentially private versions have been developed for large classes of estimators—including those used routinely by empirical economists—often with little effective cost in terms of accuracy of the released results.[15]

Multiple Statistics

Of course, researchers wish to publish more than one statistic per database. In our example above, the privacy budget ϵ was divided between two statistics, n and *sum*, and each was then independently computed in a differential-privacy preserving way.

An alternative approach is to compute the two statistics jointly. Our above definition of sensitivity generalizes to multiple (real valued) statistics by turning f from a scalar function into a vector function, and replacing the difference $|f(D) - f(D')|$ with the sum of differences component by component (known as *taxicab distance*, or L_1 *distance*). The Laplace noise technique also generalizes, to prescribe noise with standard deviation $\sqrt{2}\Delta f/\epsilon$ as before (but with the generalized Δf), drawn i.i.d. and added separately to each component of the the true f.

[15] An early survey by Dwork and Smith (2010) discusses in detail some of the first results (see also Wasserman and Zhou (2010) for an early discussion of nonparametric density estimators). Here, we briefly highlight some of the statistical objectives treated in the literature. Connections between differential privacy and robust statistics were first explored by Dwork and Lei (2009), who demonstrate differentially private algorithms for interquartile distance, median, and linear regression. Lei (2011) and Nekipelov and Yakovlev (2011) study differentially private M-estimators. Smith (2008, 2011), extending connections to robustness, finds that for almost any estimator that is asymptotically normal on i.i.d. samples from the underlying distribution (e.g., parametric MLE estimators, logistic regression, and linear regression, under regularity conditions), there are differentially private versions with asymptotically no additional perturbation. Chaudhuri and Hsu (2012) explore another aspect of robustness, showing that the convergence rate of any accurate, differentially private estimator is tied to its gross error sensitivity. Chaudhuri, Monteleoni and Sarwate (2011), Rubinstein et al. (2012), and Kifer, Smith and Thakurta (2012) provide approaches specifically tailored to minimize a convex loss function, allowing for, e.g., privacy-preserving logistic regression and support vector machines. Quite recently, Thakurta and Smith (2013) provide generic results establishing differentially private versions of stable model selection procedures. They give specific results for sparse linear regression via a new analysis of the stability of the Lasso estimator. Finally, in addition to these theoretical results, a number of papers empirically investigate the performance of differentially private estimators (see, e.g., Vu and Slavkovic, 2009; Chaudhuri, Monteleoni and Sarwate, 2011; Abowd, Schneider and Vilhuber, 2013).

This alternative approach may significantly reduce the amount of added noise, as demonstrated by the case of histograms (Dwork et al., 2006*a*). Consider the release of a frequency histogram of salaries in some database. Treating each bin as a separate statistic (e.g., "the count of rows with salary $0–10,000" is one statistic) would require dividing the privacy budget ϵ between the bins. The sensitivity (i.e., Δf) of each such bin statistic, like that of any such count query, is 1. But 1 is also the (generalized) sensitivity of the vector function consisting of the entire histogram, since adding an individual to a database always adds 1 to the count of one of the bins and 0 to all others. In this example, calculating all the bins jointly reduces the added noise because it saves the need to first divide the privacy budget between the statistics—a division whose cost in added noise increases with the number of bins. More generally, consider the maximum possible effect on a statistic of adding one individual to the database; if such worst-case effect cannot occur on each of a group of statistics at the same time, considering them jointly may improve results.

One of the main focuses of research in differential privacy in recent years has been to develop algorithms that can handle very large numbers of queries jointly with far less noise than simple noise addition would permit. This substantial literature, beginning with Blum, Ligett and Roth (2008) and continuing most recently with Hardt and Rothblum (2010) and Hardt, Ligett and McSherry (2012), develops techniques for generating "synthetic data"—a set of valid database rows—that approximate the correct answers to all of a large, fixed set of queries. The techniques go far beyond simply perturbing the data, involving ideas from geometry and computational learning theory; individual records in the resulting synthetic data are artificially generated in a sophisticated manner and cannot be connected with a single or small number of records in the original data. These approaches have started to show practicality, in the form of simple implementations that achieve good accuracy when tested on common statistical tasks using standard benchmark data (Hardt, Ligett and McSherry, 2012), but much remains to be done.[16]

From Intuitions to Provable Guarantees

What insights can the differential privacy literature offer regarding the cautionary tales above? What tools could it provide for researchers working with data? We offer some thoughts, and

[16] Another growing literature considers large sets of queries of a particular type, and aims to get a better understanding of the privacy-accuracy tradeoffs for a specific combined task. One application that has received substantial attention is contingency tables, which are computed from sets of *k-way marginal* queries; see, e.g., Barak et al. (2007), Fienberg, Rinaldo and Yang (2010), Kasiviswanathan et al. (2010), Thaler, Ullman and Vadhan (2012), Chandrasekaran et al. (2013), and Dwork, Nikolov and Talwar (2013).

highlight how different approaches respond differently to the inherent, unavoidable tradeoff between privacy and accuracy. We then discuss some of the limitations, as well as additional applications, of differential privacy.

Lessons and Reflections

In the Massachusetts GIC case—and, more generally, regarding the "anonymization" of complex data sets—lessons from differential privacy suggest considering two alternatives:

1. One could have released a differentially private, synthetic (i.e., artificial) version of the original database, after removing or coarsening complex fields such as text (which, without coarsening, would have made the data too high-dimensional for synthetization to work in practice). The synthetic data would only be useful for a pre-determined (though potentially quite large) set of statistics.

2. One could have withheld the full data but provided a differentially private interface to allow researchers (or possibly the general public) to issue queries against the database.

Both approaches—providing a sanitized database, and providing sanitized answers to individual queries—face the inescapable tradeoff between privacy and usefulness (or accuracy). To achieve privacy, they limit usefulness in different ways: while the first approach limits in advance the type of queries (and hence of analysis) possible, the second maintains flexibility but might more severely limit the overall *number* of queries, since the system has to dynamically (hence potentially less efficiently) manage a privacy budget to answer arbitrary queries as they arrive, and would eventually run out of its ϵ budget and have to refuse new queries. This idea of an overall limit—a privacy budget that places a quantifiable constraint on any approach—is a useful metaphor that highlights one of the costs of preserving privacy: it imposes fundamental limits on how much information can be revealed about the data.

In the case of the AOL debacle, the data to be released were so high-dimensional (the space of rows being all possible search histories) that they clearly could not be handled with differential privacy without some initial dimension reduction. This in itself is worth observing—free text and other high-dimensional data (e.g., genetic information) are potentially extraordinarily revealing, and deserve careful attention. Korolova et al. (2009), in response to AOL's data release, propose releasing an alternate data structure called a *query click graph*, and demonstrate on real search log data that a differentially private query click graph can be used to successfully perform some research tasks that one might typically run on search logs. As the authors note, it remains to be seen how broadly useful such sanitized data are, but such findings "offer a glimmer of hope."

Regarding the Netflix challenge, the *manner* in which it was carried out—releasing a large, very high-dimensional dataset—is unlikely to be successful under differential privacy. However, the *goals* of the challenge—namely, producing recommendations from collective user behavior—could be achievable while guaranteeing differential privacy. To explore this possibility, McSherry and Mironov (2009) evaluate several of the algorithmic approaches used in the challenge, showing that they could have been implemented in a differentially private manner (via privacy-preserving queries issued against the database) without significant impact on their accuracy.

The Facebook goal—giving advertisers a count of the number of times their ad was shown—at first sounds as if it might be well suited to differential privacy: one could simply add an appropriate level of Laplace noise to the true count. However, charging advertisers based on noisy counts may be considered objectionable, and regardless, privacy would then degrade with the number of ad campaigns (or, alternatively, Facebook would have to discontinue the service once they ran out of a certain ϵ budget they had committed to). Even if we assume that advertisers do not share the statistics Facebook reports to them (and so perhaps each advertiser can be apportioned a separate privacy budget rather than sharing a single budget among them all), large advertisers likely run so many campaigns that the noise necessary in order to ensure any reasonable level of privacy would swamp any signal in the data. Korolova (2011) suggests that an approach like differential privacy would provide the most robust starting point for privately addressing Facebook's goal, and discusses these and other challenges that leave the targeted-ads application an intriguing open problem.

More generally, what tools and other thoughts could differential privacy potentially offer to those of us who work with data?

While no standardized implementations yet exist, and while conventions (e.g., regarding setting ϵ) have not yet been established, a rich set of theoretical results already provides the foundations for a useful toolbox for the data-based researcher. If one would like to publish a single statistic (or a small set of statistics), differentially private estimator versions might already exist. As discussed above, the accuracy cost imposed by the added noise may be negligible when n is sufficiently large. Regardless of whether the statistic of interest has received attention in the differential privacy literature, the study of differential privacy suggests that it may be helpful to understand the *sensitivity* of the statistic to changes in one person's information—how much can varying one entry of the database affect the statistic? Such understanding not only helps assess how much noise one could add to achieve differential privacy in the simplest manner; it is also helpful for getting an intuitive understanding of how and why a statistic might be revealing. There are also techniques for differentially private release of statistics that may not have low sensitivity in the worst case, but are suspected to have low sensitivity on the data of interest (Nissim, Raskhodnikova and Smith, 2007; Dwork and Lei, 2009). Finally, if one wishes to publish a large set of statistics or produce sanitized data, as we discussed, general purpose techniques for doing so already exist, but it is possible that a researcher's particular properties of interest would be even better served by a specialized differentially private mechanism.

The centrality of the notion of sensitivity to work on differential privacy highlights an old truth from a new perspective: it underscores the importance of thinking about the robustness of the statistics we report. If reporting a statistic while preserving privacy requires introducing an unacceptable level of randomness, this may indicate that one's dataset is too small for one's desired levels of privacy and accuracy; but it may also suggest that worst-case scenarios exist under which the statistic is simply not robust: it may be too sensitive to potential individual outliers.

Finally, the differential privacy definition offers one way to quantify the often-loosely-used notions of *privacy* and *anonymity*. Researchers may find such quantification helpful in thinking about whether study participants should be given a different, more qualified, promise of privacy/anonymity than is standardly given—especially in settings where implementing a specific guarantee (not necessarily the one offered by differential privacy) is not practical.

Limitations

Like any other rigorous approach, the differential privacy approach makes some assumptions that may be questioned. For example, it assumes that an individual's private data are conveniently represented as a row in a database, and it implicitly assumes that a particular definition captures what we mean by privacy.

Strong privacy guarantees necessarily obscure information. The intentional introduction of randomness into published outcomes may require adjustments to specific implementations of scientific replication. More generally, for some applications, the very idea of deliberately introducing randomness is problematic: preventable mistakes such as allocating the wrong resources to the wrong individuals or making the wrong policy decisions could have grave consequences.

As hinted above, a potential limitation of differentially private mechanisms producing synthetic data is that they require the data analyst to specify her query set in advance. Many times, one may not know in advance exactly which statistics one wishes to compute or what properties of a dataset must be preserved in order for it to be useful. There is a natural tension between an analyst's desire to "look at the data" before deciding what to do with it, and a privacy researcher's desire that all computations that touch the original data be made formal and privacy-preserving. As a practical response to this limitation, rather than attempting to define the query set a priori, one could consider using some of the privacy budget for *interactive queries* where the analyst poses queries one at a time and receives privacy-preserving answers, and may then base her choice of future queries on the answer she has received so far. Once the analyst has established via this sequence of interactive queries what properties of the original database to preserve in the sanitized version, she can use the rest of her privacy budget to produce sanitized data. More generally, with the growth of big data, the "look at the data" approach is destined to change—"looking" at enormous datasets really does mean running analyses on them, and as soon as "looking at the data" has a technical meaning, one can try to enable it in a privacy-preserving manner.

Finally, for particular applications, differentially private mechanisms may not yet have been developed, or the existing technology may not enable a satisfying privacy-accuracy tradeoff. When no strong lower-bound results are known for the application of interest, these are not weaknesses of the differential privacy definition, but merely suggest that more research is needed. Even when lower bounds are known, in many cases they are not specific to differential privacy, but rather reflect that certain tasks are inherently revealing and hence may be fundamentally incompatible with privacy.

Differential Privacy and Mechanism Design

The last few years have seen a growth of interest in a number of topics at the intersection of differential privacy and economics, in particular, privacy and mechanism design; see Pai and Roth (2013) for a survey. Some of the key questions under consideration include how one might incorporate privacy considerations into utility functions and how one might model the value of privacy.[17]

From a mechanism design point of view, the differential privacy guarantee—that a participant's inclusion or removal from the database would have almost no effect on the outcome—could be viewed as a valuable guarantee even in the absence of privacy concerns. In particular, consider settings where participants in a database can misrepresent their individual data, and have pref-

[17] Work in this area includes Ghosh and Roth (2011), Nissim, Orlandi and Smorodinsky (2012), Fleischer and Lyu (2012), Roth and Schoenebeck (2012), Ligett and Roth (2012), Xiao (2013), Chen et al. (2013), and Ghosh and Ligett (2013).

erences over the possible outcomes of a function to be computed from the data. A differentially private computation implies that such participants have only limited incentive to lie: lying would have only a limited effect on the outcome. McSherry and Talwar (2007) were the first to observe this connection, i.e., that differential privacy implies asymptotic (or approximate) strategyproofness (or truthfulness). Of course, under differential privacy, not only do individuals have almost no incentive to lie; they also have almost no incentive to tell the truth (Nissim, Smorodinsky and Tennenholtz, 2012; Xiao, 2013). However, a small psychological cost of lying could strictly incentivize truth-telling.

For our purposes, in the context of data collection, analysis, and reporting, this approximate truthfulness implication may be of particular interest to researchers who wish to gather survey data in settings where participation is voluntary and the accuracy of responses cannot be easily verified. More generally, the asymptotic strategyproofness implied by differential privacy inherits some of the latter's useful additional properties, such as group privacy and composition. Hence, it provides immediate guarantees in the presence of k colluding individuals (a collusion resistance that deteriorates with the coalition size k), and it holds under repeated application of the mechanism (with the same deterioration with the number of repetitions).

Finally, this asymptotic truthfulness has inspired further work on privacy-preserving mechanism design (Huang and Kannan, 2012; Kearns et al., 2012) and has enabled differential privacy to be used as a tool in the design of truly strategyproof mechanisms (see, e.g., Nissim, Smorodinsky and Tennenholtz, 2012).

Concluding Thoughts

Privacy concerns in the face of unprecedented access to big data are nothing new. More than thirty-five years ago, Dalenius (1977) already discusses "the proliferation of computerized information system[s]" and "the present era of public concern about 'invasion of privacy'." But as big data get bigger, so do the concerns. Greely (2007) discusses genomic databases, concluding that

> [t]he size, the cost, the breadth, the desired broad researcher access, and the likely high public profile of genomic databases will make these issues especially important to them. Dealing with these issues will be both intellectually and politically difficult, time-consuming, inconvenient, and possibly expensive. But it is not a solution to say that "anonymity" means only "not terribly easy to identify," ... or that "informed consent"

is satisfied by largely ignorant blanket permission.

Replacing "genomic databases" with "big data" in general, our overall conclusion may be similar. The stories in the first part of this paper demonstrate that relying on intuition when attempting to protect subject privacy may not be enough. Moreover, privacy failures may occur even when the raw data are never publicly released and only some seemingly innocuous *function* of the data, such as a statistic, is published. The purpose of these stories is to increase awareness.

The ideas from the differential privacy literature we introduce in the second part of this paper provide one formal way for thinking about the notion of privacy that researchers may want to guarantee to subjects. They also provide a framework, or a tool, for thinking quantitatively about privacy-accuracy tradeoffs. We would like to see more such thinking among data-based researchers. In particular, with computer scientists using phrases such as "the amount of privacy loss" and "the privacy budget," the time seems ripe for more economists to join the conversation. Is a certain lifetime amount of ϵ a basic right? Is privacy a term in the utility function that can in principle be compared against the utility from access to accurate data? Should fungible, transferable ϵ be allowed to be sold in markets from private individuals to potential data users, and if so, what would its price be? Should a certain privacy budget be allocated across interested users of publicly owned (e.g., Census) data, and if so, how? Such questions are beginning to receive attention, as mentioned above. Increased attention may eventually bring change to common practices.

What kind of changes could one envision? In the third part of our paper we discuss specific applications of differential privacy to concrete situations, highlighting some limitations. When big data means large n, an increasing number of common computations can be achieved in a differentially private manner, with little cost to precision. It is not inconceivable that within a few years, many of the computations that have been—and those that are yet to be—proven achievable in theory, will be applied in practice. Echoing Dwork and Smith (2010), who "would like to see a library of differentially private versions of the algorithms in R and SAS," we would be happy to have a differentially private option in estimation commands in STATA. But ready-to-use, commercial-grade applications will not be developed without sufficient demand from potential users. We hope that the incorporation of privacy considerations into the vocabulary of empirical researchers will help raise demand, and stimulate further discussion and research—including, we hope, regarding additional approaches to privacy.

Until such applications are available, it might be wise to pause and reconsider researchers' promises and, more generally, obligations to subjects. When researchers (and IRBs!) are confident that the data pose only negligible privacy risks—e.g., some innocuous small surveys and lab experiments—it may be preferable to replace promises of anonymity with promises for "not terribly easy" identification or, indeed, with no promises at all. In particular, researchers could explicitly inform consenting subjects that a determined attacker may be able to identify them in posted data, or even learn things about them merely by looking at the empirical results of a research paper. We caution against taking the naive alternate route of simply refraining from making harmless data publicly available; freedom of information, access to data, transparency, and scientific replication are all dear to us.[18] Of course, the tradeoffs, and in particular the question of what privacy risks are negligible and what data are harmless, should be carefully considered and discussed; a useful question to ask ourselves may resemble a version of the old newspaper test: would our subjects mind if their data were identified and published in the *New York Times*?

References

Abowd, John M., Matthew J. Schneider, and Lars Vilhuber. 2013. "Differential Privacy Applications to Bayesian and Linear Mixed Model Estimation." *Journal of Privacy and Confidentiality*, 5(1).

Barak, Boaz, Kamalika Chaudhuri, Cynthia Dwork, Satyen Kale, Frank McSherry, and Kunal Talwar. 2007. "Privacy, accuracy, and consistency too: a holistic solution to contingency table release." In *Symposium on Principles of Database Systems*. 273–282.

Barth-Jones, Daniel C. 2012. "The 'Re-Identification' of Governor William Weld's Medical Information: A Critical Re-Examination of Health Data Identification Risks and Privacy Protections, Then and Now." `http://ssrn.com/abstract=2076397`.

Blum, Avrim, Katrina Ligett, and Aaron Roth. 2008. "A learning theory approach to non-interactive database privacy." In *Symposium on the Theory of Computing*. 609–618.

Chandrasekaran, Karthekeyan, Justin Thaler, Jonathan Ullman, and Andrew Wan. 2013. "Faster Private Release of Marginals on Small Databases." `http://arxiv.org/pdf/1304.3754v1.pdf`.

[18] Flood et al. (2013) provide a comprehensive discussion of such a transparency-confidentiality tradeoff in a context that is very different from ours, yet of great interest to economists—that of financial supervision and regulation.

Chaudhuri, Kamalika, and Daniel J. Hsu. 2012. "Convergence Rates for Differentially Private Statistical Estimation." In *International Conference on Machine Learning.* 1327–1334.

Chaudhuri, Kamalika, Claire Monteleoni, and Anand D. Sarwate. 2011. "Differentially private empirical risk minimization." *Journal of Machine Learning Research*, 12: 1069–1109.

Chen, Yiling, Stephen Chong, Ian A. Kash, Tal Moran, and Salil P. Vadhan. 2013. "Truthful mechanisms for agents that value privacy." In *Conference on Electronic Commerce.* 215–232.

Dalenius, Tore. 1977. "Towards a methodology for statistical disclosure control." *Statistisk tidskrift*, 15: 429–444.

Dinur, Irit, and Kobbi Nissim. 2003. "Revealing information while preserving privacy." In *Symposium on Principles of Database Systems.* 202–210.

Dwork, Cynthia. 2006. "Differential privacy." In *International Conference on Automata, Languages and Programming.* 1–12.

Dwork, Cynthia. 2011*a*. "A firm foundation for private data analysis." *Communications of the ACM*, 54(1): 86–95.

Dwork, Cynthia. 2011*b*. "The promise of differential privacy: A tutorial on algorithmic techniques." In *Foundations of Computer Science.* 1–2.

Dwork, Cynthia, Aleksandar Nikolov, and Kunal Talwar. 2013. "Efficient Algorithms for Privately Releasing Marginals via Convex Relaxations." `http://arxiv.org/pdf/1308.1385v1.pdf`.

Dwork, Cynthia, and Adam Smith. 2010. "Differential privacy for statistics: What we know and what we want to learn." *Journal of Privacy and Confidentiality*, 1(2): 135–154.

Dwork, Cynthia, and Jing Lei. 2009. "Differential privacy and robust statistics." In *Symposium on Theory of Computing.* 371–380.

Dwork, Cynthia, and Kobbi Nissim. 2004. "Privacy-preserving datamining on vertically partitioned databases." In *International Conference on Cryptology (CRYPTO).* 528–544.

Dwork, Cynthia, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006*a*. "Calibrating noise to sensitivity in private data analysis." In *Theory of Cryptography Conference*. 265–284.

Dwork, Cynthia, Frank McSherry, Kobbi Nissim, and Adam Smith. 2011. "Differential Privacy: A Primer for the Perplexed." In *Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality*. `http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2011/26_Dwork-Smith.pdf`.

Dwork, Cynthia, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. 2006*b*. "Our Data, Ourselves: Privacy Via Distributed Noise Generation." In *International Conference on the Theory and Applications of Cryptographic Techniques (EUROCRYPT)*. 486–503.

Fienberg, Stephen E., Alessandro Rinaldo, and Xiaolin Yang. 2010. "Differential privacy and the risk-utility tradeoff for multi-dimensional contingency tables." In *Privacy in Statistical Databases*. 187–199.

Fleischer, Lisa, and Yu-Han Lyu. 2012. "Approximately Optimal Auctions for Selling Privacy when Costs are Correlated with Data." In *Conference on Electronic Commerce*. 568–585.

Flood, Mark, Jonathan Katz, Stephen Ong, and Adam Smith. 2013. "Cryptography and the Economics of Supervisory Information: Balancing Transparency and Condentiality." Unpublished.

Ghosh, Arpita, and Aaron Roth. 2011. "Selling privacy at auction." In *Conference on Electronic Commerce*. 199–208.

Ghosh, Arpita, and Katrina Ligett. 2013. "Privacy and Coordination: Computing on Databases with Endogenous Participation." In *Conference on Electronic Commerce*. 543–560.

Golle, Philippe. 2006. "Revisiting the uniqueness of simple demographics in the US population." In *Workshop on Privacy in Electronic Society*. 77–80.

Greely, Henry T. 2007. "The uneasy ethical and legal underpinnings of large-scale genomic biobanks." *Annual Review of Genomics and Human Genetics*, 8: 343–364.

Hardt, Moritz, and Guy N. Rothblum. 2010. "A Multiplicative Weights Mechanism for Privacy-Preserving Data Analysis." In *Foundations of Computer Science*. 61–70.

Hardt, Moritz, Katrina Ligett, and Frank McSherry. 2012. "A Simple and Practical Algorithm for Differentially Private Data Release." In *Advances in Neural Information Processing Systems.* 2348–2356.

Huang, Zhiyi, and Sampath Kannan. 2012. "The Exponential Mechanism for Social Welfare: Private, Truthful, and Nearly Optimal." In *Foundations of Computer Science.* 140–149.

Kasiviswanathan, Shiva Prasad, and Adam Smith. 2008. "A Note on Differential Privacy: Defining Resistance to Arbitrary Side Information." `http://arxiv.org/pdf/0803.3946v1.pdf`.

Kasiviswanathan, Shiva Prasad, Mark Rudelson, Adam Smith, and Jonathan Ullman. 2010. "The price of privately releasing contingency tables and the spectra of random matrices with correlated rows." In *Symposium on the Theory of Computing.* 775–784.

Kearns, Michael, Mallesh Pai, Aaron Roth, and Jon Ullman. 2012. "Mechanism Design in Large Games: Incentives and Privacy." `http://arxiv.org/abs/1207.4084`.

Kifer, Daniel, Adam Smith, and Abhradeep Thakurta. 2012. "Private convex empirical risk minimization and high-dimensional regression." In *Conference on Learning Theory.* 25.1–25.40.

Klarreich, Erica. 2012. "Privacy by the Numbers: A New Approach to Safeguarding Data." *Quanta Magazine.* `https://www.simonsfoundation.org/quanta/20121210-privacy-by-the-numbers-a-new-approach-to-safeguarding-data/`.

Korolova, Aleksandra. 2011. "Privacy Violations Using Microtargeted Ads: A Case Study." *Journal of Privacy and Confidentiality*, 3(1).

Korolova, Aleksandra, Krishnaram Kenthapadi, Nina Mishra, and Alexandros Ntoulas. 2009. "Releasing search queries and clicks privately." In *Proceedings of the 18th international conference on World wide web.* 171–180.

Kumar, Ravi, Jasmine Novak, Bo Pang, and Andrew Tomkins. 2007. "On anonymizing query logs via token-based hashing." In *Conference on the World Wide Web.* 629–638.

Lei, Jing. 2011. "Differentially private m-estimators." In *Advances in Neural Information Processing Systems.* 361–369.

Ligett, Katrina, and Aaron Roth. 2012. "Take it or Leave it: Running a Survey when Privacy Comes at a Cost." In *Workshop on Internet and Network Economics (WINE).* 378–391.

Machanavajjhala, Ashwin, Daniel Kifer, John M. Abowd, Johannes Gehrke, and Lars Vilhuber. 2008. "Privacy: Theory meets Practice on the Map." In *International Conference on Data Engineering.* 277–286.

McSherry, Frank, and Ilya Mironov. 2009. "Differentially private recommender systems: building privacy into the net." In *Conference on Knowledge Discovery and Data Mining.* 627–636.

McSherry, Frank, and Kunal Talwar. 2007. "Mechanism Design via Differential Privacy." In *Symposium on Foundations of Computer Science (FOCS).* 94–103.

Narayanan, Arvind, and Vitaly Shmatikov. 2008. "Robust de-anonymization of large sparse datasets." In *Symposium on Security and Privacy.* 111–125.

Nekipelov, Denis, and Evegeny Yakovlev. 2011. "Private extremum estimation." `http://emlab.berkeley.edu/~nekipelov/pdf_papers/paper3.pdf`.

Nissim, Kobbi, Claudio Orlandi, and Rann Smorodinsky. 2012. "Privacy-aware mechanism design." In *Conference on Electronic Commerce.* 774–789.

Nissim, Kobbi, Rann Smorodinsky, and Moshe Tennenholtz. 2012. "Approximately optimal mechanism design via differential privacy." In *Innovations in Theoretical Computer Science Conference.* 203–213.

Nissim, Kobbi, Sofya Raskhodnikova, and Adam Smith. 2007. "Smooth sensitivity and sampling in private data analysis." In *Symposium on Theory of Computing.* 75–84.

Ohm, Paul. 2010. "Broken promises of privacy: Responding to the surprising failure of anonymization." *UCLA Law Review*, 57: 1701–1777.

Pai, Mallesh, and Aaron Roth. 2013. "Privacy and Mechanism Design." *Sigecom Exchanges*, 12(1).

Roth, Aaron, and Grant Schoenebeck. 2012. "Conducting Truthful Surveys, Cheaply." In *Conference on Electronic Commerce.* 826–843.

Rousseau, Peter L. 2013. "Report of the Secretary." *American Economic Review*, 103(3): 669–72.

Rubinstein, Benjamin I.P., Peter L. Bartlett, Ling Huang, and Nina Taft. 2012. "Learning in a Large Function Space: Privacy-Preserving Mechanisms for SVM Learning." *Journal of Privacy and Confidentiality*, 4(1): 65–100.

Smith, Adam. 2008. “Efficient, differentially private point estimators.” `http://arxiv.org/pdf/0809.4794v1.pdf`.

Smith, Adam. 2011. “Privacy-preserving statistical estimation with optimal convergence rates.” In *Symposium on Theory of Computing*. 813–822.

Sweeney, Latanya. 1997. “Weaving technology and policy together to maintain confidentiality.” *The Journal of Law, Medicine & Ethics*, 25(2-3): 98–110.

Sweeney, Latanya. 2002. “Achieving k-anonymity privacy protection using generalization and suppression.” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05): 571–588.

Sweeney, Latanya. 2013. “Matching Known Patients to Health Records in Washington State Data.” `http://thedatamap.org/1089-1.pdf`.

Sweeney, Latanya, Akua Abu, and Julia Winn. 2013. “Identifying Participants in the Personal Genome Project by Name.” `http://dataprivacylab.org/projects/pgp/1021-1.pdf`.

Thakurta, Abhradeep Guha, and Adam Smith. 2013. “Differentially Private Feature Selection via Stability Arguments, and the Robustness of the Lasso.” In *Conference on Learning Theory*. 819–850.

Thaler, Justin, Jonathan Ullman, and Salil Vadhan. 2012. “Faster algorithms for privately releasing marginals.” In *Automata, Languages, and Programming*. 810–821.

Vu, Duy, and Aleksandra Slavkovic. 2009. “Differential privacy for clinical trial data: Preliminary evaluations.” In *Conference on Data Mining Workshops*. 138–143.

Wasserman, Larry, and Shuheng Zhou. 2010. “A statistical framework for differential privacy.” *Journal of the American Statistical Association*, 105(489): 375–389.

Xiao, David. 2013. “Is privacy compatible with truthfulness?” In *Innovations in Theoretical Computer Science (ITCS)*. 67–86.