An Economic Analysis of Privacy Protection and Statistical Accuracy as Social Choices

John M. Abowd and Ian M. Schmutte

August 15, 2018

Forthcoming in *American Economic Review*

Abowd: U.S. Census Bureau HQ 8H120, 4600 Silver Hill Rd., Washington, DC 20233, and Cornell University, (email: john.maron.abowd@census.gov); Schmutte: Department of Economics, University of Georgia, B408 Amos Hall, Athens, GA 30602 (email: schmutte@uga.edu). Abowd and Schmutte acknowledge the support of Alfred P. Sloan Foundation Grant G-2015-13903 and NSF Grant SES-1131848. Abowd acknowledges direct support from NSF Grants BCS-0941226, TC-1012593. Any opinions and conclusions are those of the authors and do not represent the views of the Census Bureau, NSF, or the Sloan Foundation. We thank the Center for Labor Economics at UC–Berkeley and Isaac Newton Institute for Mathematical Sciences, Cambridge (EPSRC grant no. EP/K032208/1) for support and hospitality. We are extremely grateful for very valuable comments and guidance from the editor, Pinelopi Goldberg, and six anonymous referees. We acknowledge helpful comments from Robin Bachman, Nick Bloom, Larry Blume, David Card, Michael Castro, Jennifer Childs, Melissa Creech, Cynthia Dwork, Casey Eggleston, John Eltinge, Stephen Fienberg, Mark Kutzbach, Ron Jarmin, Christa Jones, Dan Kifer, Ashwin Machanavajjhala, Frank McSherry, Gerome Miklau, Kobbi Nissim, Paul Oyer, Mallesh Pai, Jerry Reiter, Eric Slud, Adam Smith, Bruce Spencer, Sara Sullivan, Salil Vadhan, Lars Vilhuber, Glen Weyl, and Nellie Zhao along with seminar and conference participants at the U.S. Census Bureau, Cornell, CREST, George Mason, Georgetown, Microsoft Research–NYC, University of Washington Evans School, and SOLE. William Sexton provided excellent research assistance. No confidential data were used in this paper. Supplemental materials available at http://doi.org/10.5281/zenodo.1345775. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

Abstract

Statistical agencies face a dual mandate to publish accurate statistics while protecting respondent privacy. Increasing privacy protection requires decreased accuracy. Recognizing this as a resource allocation problem, we propose an economic solution: operate where the marginal cost of increasing privacy equals the marginal benefit. Our model of production, from computer science, assumes data are published using an efficient differentially private algorithm. Optimal choice weighs the demand for accurate statistics against the demand for privacy. Examples from U.S. statistical programs show how our framework can guide decision-making. Further progress requires a better understanding of willingness-to-pay for privacy and statistical accuracy.

National statistical agencies collect information about the population and economy of a country directly from its people and businesses. They face a dual mandate to publish useful summaries of these data while protecting the confidentiality of the underlying responses. These mandates are enshrined in law and official practices.[1] For example, the Census Bureau is required by Article I of the U.S. Constitution to enumerate the population every ten years and by legislation to publish data from that census for the purpose of redrawing every legislative district in the country.[2] When providing these data, the Census Bureau is also subject to a legal prohibition against "mak[ing] any publication whereby the data furnished by any particular establishment or individual ... can be identified."[3]

The fundamental challenge posed in servicing this dual mandate is that as more statistics are published with more accuracy, more privacy is lost (Dinur and Nissim 2003). Economists recognize this as a problem of resource allocation. We propose an economic framework for solving it. Statistical agencies must allocate the information in their collected data between two competing uses: production of statistics that are sufficiently accurate balanced against the protection of privacy for those in the data. We combine the economic theory of public goods with cryptographic methods from computer science to show that social welfare maximization can, and should, guide how statistical agencies manage this trade-off.

Figure 1 illustrates our adaptation of the approach proposed by Samuelson (1954) for the efficient allocation of public goods. The horizontal axis measures *privacy loss* parameterized by ε. The vertical axis measures *accuracy*, parameterized by I. We define both concepts in detail in Section II. The line labeled *PF* represents the production function, which describes feasible combinations of privacy loss and statistical accuracy available to the agency, given its endowment of data and known mechanisms for publishing. The line labeled *SWF* is an indifference curve from the social welfare function defined in Section IV. It describes

[1] The Confidential Information Protection and Statistical Efficiency Act (CIPSEA) (44 U.S. Code 2002) and Census Act (13 U.S. Code 1954) obligate U.S. statistical agencies to protect confidentiality.

[2] Under Public Law 94-171.

[3] 13 U.S. Code (1954, Section 9.a.2.).

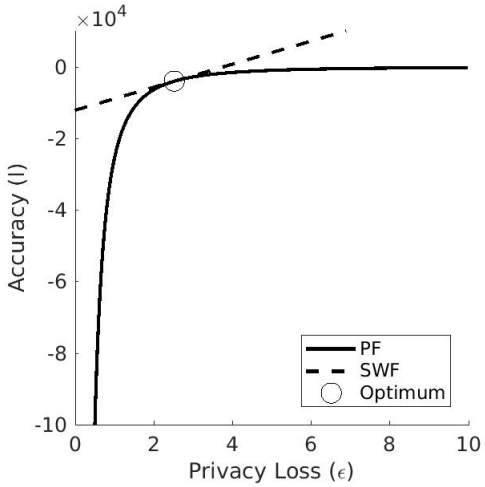

Figure 1: Solution to the Planner's Problem

aggregate preferences for privacy loss and accuracy. The optimal combination of privacy loss and accuracy is indicated by the open circle. At that point, the marginal rate of transformation, which measures the cost of increased loss of privacy, matches the social willingness to accept privacy loss, both measured in units of increased statistical accuracy.

To date, this social choice framework has not been adapted to help guide statistical agencies in fulfilling their dual mandate. Our key insight is that formal privacy systems developed in computer science can characterize the levels of privacy and accuracy available to a data custodian as a production function. In our model, a statistical agency uses a known *differentially private* mechanism to publish official statistics (Cynthia Dwork, Frank McSherry, Kobbi Nissim and Adam Smith 2006; Cynthia Dwork 2008; Cynthia Dwork and Aaron Roth 2014). Differential privacy measures privacy loss by the amount of information about any data record that is leaked when statistics are published. Differential privacy is very useful for describing the production technologies available to the statistical agency. However, formal privacy models shed no light on how to choose the right level of privacy or accuracy. This social choice question requires the tools of economics.[4]

[4]In related contributions from the electronic commerce literature, Hsu et al. (2014) and Ghosh

Although the approach outlined in Figure 1 is familiar to economists, threats to privacy from publishing summary statistics may not be. Before discussing technical details, we discuss in Section I how differential privacy relates to practical, real-world concerns about confidentiality and data security. We give a plain-language interpretation of differential privacy and its relationship to the concerns individuals have about protecting whatever sensitive information has been stored in a database. We also discuss specific data breaches to highlight why statistical agencies and private data custodians are rapidly embracing differential privacy and other formal privacy-preserving data publication systems.

We model the publication of population statistics in Sections II and III. Section II focuses on the concepts we borrow from computer science. In Section III, we interpret differentially-private publication mechanisms through the lens of producer theory. We give examples of mechanisms that yield closed, bounded, and convex production functions relating privacy loss and statistical accuracy. We define statistical accuracy as the expected squared error between the published value and the value that would be published in the absence of privacy protection. In Section IV, we model preferences for privacy protection and accuracy as public goods and describe their optimal levels using a utilitarian social welfare function.

The issues raised in this paper are far from academic. The threats to privacy inherent in the "big data" era have affected the policies governing statistical agencies. In September 2017, the bi-partisan Commission on Evidence-based Policymaking recommended that agencies "[a]dopt modern privacy-enhancing technologies ... to ensure that government's capabilities to keep data secure and protect confidentiality are constantly improving" (Commission on Evidence-Based

and Roth (2015) model market-based provision of statistical summaries for private use by a single analyst in a setting where individuals can be directly compensated for their associated loss of privacy. In the context of these private transactions, Hsu et al. (2014) characterize an economic approach to setting the level of privacy loss. These models, like the formal privacy literature more generally, do not address the public-good nature of the published statistics and privacy protection offered by statistical agencies. The fundamental difference in our paper is the development of a social choice framework that accounts for all the benefits from the published statistics, not just those accruing to the explicit players as in the Hsu et al. and Ghosh and Roth models, and for all privacy-loss costs, not just the explicit losses for those who opt-in, because participation in our data collection is mandatory.

Policymaking 2017, p. 2). That same month, the U.S. Census Bureau announced that it will use differential privacy, the leading privacy-enhancing technology, to protect the publications from its 2018 End-to-End Census Test and the 2020 Census.[5] Our goal is to develop a principled framework for practical, empirically-driven, policy guidance regarding the balance of privacy protection and accuracy in modern statistical systems.

In Section V, we present an illustrative analysis of our framework applied to the allocation of federal funding to school districts under Title 1. In Section VI, we interpret several real-world problems in the U.S. statistical system using our social choice framework: the use of the PL94-171 tabulations to draw new legislative districts, the use of published inputs from the economic censuses to benchmark national accounts, the use of tax return data for tax policy modeling, and the publication of general-purpose micro-data. In each case, we describe the consequences of altering the weight on statistical accuracy versus privacy loss when facing an efficient frontier constraining the choices.

The allocation problem we study requires the perspective of economists—both in developing new theory and in connecting that theory to empirically-driven policy analysis. Until now, our discipline has ceded one of the most important debates of the information age to computer science.[6] In our conclusion, we draw attention to open questions that we hope will interest and inspire contributions from economists and other social scientists.

[5] See https://www.census.gov/about/cac/sac/meetings/2017-09-meeting.html (cited on March 12, 2018).

[6] Privacy-preserving data analysis is barely known outside of computer science. A search for "differential privacy" in JSTOR's complete economics collection through December 2017 found five articles. The same query for statistics journals found six. A search of the ACM digital library, the repository for the vast majority of refereed conference proceedings in computer science, for the same quoted keyword found 47,100 results.

I Key Concepts: Privacy and Accuracy

Defining privacy in a rigorous but meaningful way is particularly challenging. To this end, we work with the concept of *differential privacy*. This section explains how differential privacy relates to both data security and individual privacy. We also discuss our decision to measure statistical accuracy by expected squared error loss, and why this choice is not without loss of generality.

I.A Measuring Privacy

Differential privacy is a property of algorithms used to publish statistics from a confidential database. A differentially private algorithm guarantees that the published statistics will not change "too much" whether any observation from the confidential data is included or excluded. The notion of "too much" is quantified by a parameter, ε, which measures the maximum difference in the log odds of observing any statistic across similar databases. For details, see Definition 3.

Differential privacy provides provable limits on an attacker's ability to re-identify individual records based on published statistics. Other methods of confidentiality protection are vulnerable to re-identification of large numbers of records through database reconstruction. We discuss database reconstruction, re-identification attacks, and the emergence of differential privacy in Section I.B.2.

Differential privacy can also guarantee individuals that their personal information—secrets—will not be blatantly disclosed. Following computer science, we use the term *semantic privacy* to refer to the latter sense of privacy protection. Semantic privacy insures that what an attacker can learn about any person from published statistics does not depend "too much" on whether their data were used to compute the statistics.[7] In Section IV.A we describe the conditions under which publication mechanisms guaranteeing ε-differential privacy also guarantee ε-semantic privacy.

Publication systems that are differentially private are *closed under composition*,

[7] Statisticians will recognize the concept of semantic privacy as directly related to inferential disclosure (Dalenius 1977; Goldwasser and Micali 1982).

meaning that the cumulative privacy loss incurred by running multiple analyses on the same database can be computed from the privacy loss bounds on the component algorithms. Differentially private systems are *robust to post-processing*, meaning that the privacy guarantee cannot be compromised by manipulation of the outputs, even in the presence of arbitrary outside information. Differentially private systems are *future proof*, meaning that their privacy guarantees do not degrade as technology improves or new data from other sources are published. Finally, differential privacy systems are *public*, meaning that all the algorithms and parameters, except for the random numbers used in the implementation, can be published without compromising the privacy guarantee. As argued in Abowd and Schmutte (2015), the public property of differential privacy is a major advantage over traditional statistical disclosure limitation (SDL) because it allows for verification of the privacy protection and correct inferences from the published statistics.

There are caveats. While the differential privacy bound is useful for characterizing production possibilities, its interpretation is different from other economic variables that are conceptually well-defined but not precisely measurable. Differential privacy is a nominal bound on the worst-case privacy loss faced by any individual. Realized privacy loss depends on the actual data, the published statistics, and external information that is, or may become, available. We characterize the statistical agency as choosing how to operate a system that guarantees a bound on everyone's privacy loss of ε, and allows verifiable accounting of compliance with that guarantee.

The worst-case bound in the definition of differential privacy is necessary for closure under composition. Closure under composition allows the custodian to compute global privacy loss as ε for the entire set of published statistics. Closure under composition also preserves the non-rival public-good property of differential privacy protection. The global privacy protection parameterized by ε is therefore the relevant public good that an economic analysis must allocate, not the loss of privacy experienced by a particular individual after publication of a particular statistic.

I.B Data Security and Privacy

It may not be immediately obvious to applied researchers how publishing summary data can threaten privacy. To motivate our analysis, we describe real-world violations of data security and situate them within a framework that reflects the common understanding of confidentiality protection in economics, statistics and computer science. We defend the premise that blatant breaches of data security—"hacks"—are unacceptable. The key insight from computer science is that publication of summary statistics leaks the same kind of private information as a breach. Differential privacy has emerged as a focal paradigm because it can provably circumscribe such leakages.

I.B.1 Motivating Examples

For an example of the harms of breaching data security, one need look no further than the Census Bureau's activities in WWII, releasing small-area data for the purposes of Japanese internment to the Army and providing individual records of the Japanese Americans in Washington, DC to the Secret Service for purposes of surveillance (Jones 2017). Statutory barriers now prevent explicit data-sharing of this sort, and the U.S. Census Bureau staunchly guards those barriers (U.S. Census Bureau 2002). The Secretary of Commerce's March 26, 2018 direction to include a question on citizenship on the 2020 Census in support of providing block-level data on the citizen voting-age population by race and ethnicity makes the question of how best to protect the confidentiality of the micro-data in these publications even more salient.[8] There remains a threat that detailed data on a sensitive population could be accidentally shared by publishing so much summary information that the underlying data can be reverse-engineered. Therefore, statistical publications should guard against *database reconstruction*.

Privacy is also threatened by *data re-identification*. In 2006, Netflix ran a contest to improve its recommendation system. Netflix released data from its database of the ratings histories of its subscribers. To protect their users' privacy, Netflix

[8] U.S. Department of Commerce (2018).

removed direct identifiers and then released only a random sample of ratings. These efforts were not sufficient. Narayanan and Shmatikov (2008) successfully re-identified a large share of the users in the Netflix Prize data by probabilistically matching to data scraped from `IMDb.com`, a crowd-sourced movie review site. This attack harmed the re-identified users because the attacker learned about movies they had rated in Netflix that they had not publicly rated in IMDb.[9]

I.B.2 From Data Breaches to Differential Privacy

To understand why differential privacy has captured the attention of statisticians, computer scientists, and economists, we elaborate on the concepts of *database reconstruction* and *data re-identification*.[10] These concepts summarize key ideas from the literatures on formal privacy and SDL, a complete review of which is beyond the scope of this paper.[11]

A *reconstruction attack* is an attempt to build a record-by-record copy of a confidential database using only statistics published from it. This record-level image reproduces all the variables that were used in any published statistic to some level of accuracy. The accuracy depends on how many linearly independent statistics were published. Call the list of variables subject to reconstruction "list A". Most published statistics do not include exact identifiers like name, address, Social Security Number (SSN), employer identification number, or medical case identifier. Call this list of identifiers "list B".

A *re-identification attack* involves the linkage of records containing variables on list B to records with variables on list A, either deterministically or probabilistically. Records that link are called *putative re-identifications*. They are unconfirmed claims that some entity on list A belongs to a specific record associated with variables on list B. A reconstruction attack abets a re-identification attack by facil-

[9] Garfinkel (2015) provides a more exhaustive overview of re-identification attacks.

[10] We prefer the term *statistical disclosure limitation* to *anonymization* or *de-identification*. All three are synonymous. We prefer the term *re-identification* to *de-anonymization*, but they also are synonymous.

[11] See Duncan, Elliot and Salazar-González (2011) for a comprehensive review of the SDL literature. Heffetz and Ligett (2014) summarize differential privacy for economists.

itating the linkage of external information from list B. It allows the attacker to compare the data on the reconstructed variables from list A with similar variables in external databases that also contain some variables from list B, generating putative re-identifications.

Our use of differential privacy is motivated by the *database reconstruction theorem* due to Dinur and Nissim (2003), which proves that conventional SDL is inherently non-private. If the data custodian publishes many linearly independent statistics, then the confidential database can be reconstructed up to a very small error. Reconstructing the confidential variables is a data breach. Methods vulnerable to database reconstruction attacks are called *blatantly non-private*.[12]

The database reconstruction theorem sounded the death knell for the SDL methods typically used by statistical agencies. Statistical agencies consider correct re-identification to be inherently problematic (Harris-Kojetin et al. 2005, p. 103). Re-identification attacks can be facilitated by using the information from successful reconstruction attacks on public-use tables and micro-data. Meanwhile, the amount of auxiliary information available to validate re-identification is rapidly increasing (Garfinkel 2015). So is the computing power and algorithmic sophistication needed to carry out these attacks.

I.B.3 Reconstruction and Re-identification in Economic Data

We consider several examples of database reconstruction and re-identification risks in economic data. One such breach occurred when the Continuous Work History Sample (CWHS) was released to researchers. Its designers were so explicit about how the digits in the SSN were used to construct the sample that a researcher could partially reconstruct valid SSNs known to be in the released

[12]Legally, whether a reconstruction-abetted re-identification attack constitutes an actionable data breach depends upon details of the attack. The U.S. government defines a data breach with reference to personally identifiable information (Donovan 2017, p. 8) as "information that can be used to distinguish or trace an individual's identity, either alone or when combined with other information that is linked or linkable to a specific individual." The actionable consequences of a breach by this definition are implemented in agency policies (e.g. Data Stewardship Executive Policy Committee 2014).

data (Perlman and Mandel 1944; Mandel 1950; Perlman 1951; Smith 1989). The researcher could then re-identify individuals using the other variables. The data were deemed to be in violation of the Tax Reform Act of 1976, and the CWHS files were recalled (Buckler 1988, p. 56).[13]

Researchers at the Census Bureau have acknowledged that their tabular publications may be vulnerable to database reconstruction attacks (Abowd 2016; U.S. Census Bureau 2017; Abowd 2017). This recognition is based, in part, on careful analysis of the statistics in all of the publications from the 2010 Census. More than 7.7 billion unique statistical summaries were published from 309 million persons—25 statistics per person. Each linearly independent statistic is one equation in a system that can be used for a record-level database reconstruction. For more details, see Appendix A.

The county tables in the Quarterly Census of Employment and Wages (QCEW) released by the Bureau of Labor Statistics (BLS) are also vulnerable to a reconstruction attack.[14] The BLS uses primary and complementary suppression to protect these tables (Harris-Kojetin et al. 2005, p. 47), but the published summaries are exactly equal to the summaries from the confidential data. The suppressed cells can be reconstructed with great precision using the time series of county tables and known relationships among the cells. Since many of the suppressed cells contain just one or two business establishments, the reconstruction attack exposes those businesses to re-identification of their payroll and employment data.

Our last example is a database reconstruction attack that always produces exact re-identification. Genome-wide association studies (GWAS) publish the marginal distributions of hundreds of thousands of alleles from study populations that are diagnosed with the same disease. It is possible to determine if a single genome was used to construct the GWAS with very high precision (Homer et al. 2008; Yu et al. 2014; Dwork et al. 2015). An attacker who determines that a genome is in the GWAS learns that it is associated with the diagnosis defining the study popula-

[13]These are the data used by Topel and Ward (1992).

[14]Scott H. Holan, Daniell Toth, Marco A. R. Ferreira and Alan F. Karr (2010) published this attack. Toth is a BLS statistician.

tion. This is a reconstruction attack because it establishes exact genomes that were input records for the GWAS. It is a re-identification attack because the attacker learns the medical case identifier associated with the genome. That association is unique due to human biology unless the person is an identical twin.[15]

I.C The Choice of an Accuracy Measure

We define statistical accuracy in terms of expected squared-error loss (Definition 4). One might hope for an analysis where the trade-off between privacy and accuracy is independent of how the statistics are used (for instance, independent of the prior beliefs of a Bayesian receiver). In that case, for any two publications, all consumers would agree about the difference in published accuracy. Brenner and Nissim (2014) show that such universal mechanisms are impossible except in the special case of publishing a single counting query (Ghosh, Roughgarden and Sundararajan 2012). In Section III, we characterize several publication mechanisms that implicitly define a functional relationship between privacy loss and squared-error loss. We will also revisit the question of productive efficiency. Before doing so, we provide the formal, technical definitions of the concepts laid out in this section.

II Model Preliminaries

We model a trusted custodian—the statistical agency—that controls a database from which it must publish population statistics. We represent the database as a matrix with known properties, and the published statistics, called queries, as functions of the data matrix. These queries represent statistics such as contingency tables and other standard public-use products.

[15] National Institutes of Health (2014) revised the NIH rules for sharing genomic data, even when de-identified in a GWAS, to require active consent: "[t]he final GDS [Genome Data Sharing] Policy permits unrestricted access to de-identified data, but only if participants have explicitly consented to sharing their data through unrestricted-access mechanisms."

II.A The Database

The population database is a matrix, D. Each row of D contains information for one of N individuals, and each column records a separate variable or feature. D is a multi-set with rows selected from a discrete, finite-valued *data domain*, χ.[16] We denote by $|\chi|$ the cardinality of χ. This setup is very general. It can handle missing data, non-response, skip patterns, alternative sources, unique identifiers, and other complex features of real data.

II.A.1 Representation of the Database as a Histogram

The *histogram representation* of D is a $|\chi| \times 1$ vector, $x \in \mathbb{Z}^{*|\chi|}$, where $\mathbb{Z}^*$ is the set of non-negative integers. The histogram records the frequency of each feasible combination of attributes in D. For each element of the data domain, $k \in \chi$, x_k is the number of records in D with attribute combination k. The ordering of $k \in \chi$ is fixed and known, but arbitrary.

The ℓ_1 norm of x is $\|x\|_1 = \sum_{i=1}^{|\chi|} |x_i| = N$, the number of records in the database. Given two histograms x and y, $\|x - y\|_1$ measures the number of records that differ between x and y. *Adjacent histograms* are those for which the ℓ_1 distance is 1.[17]

II.A.2 Population Statistics Are Database Queries

Population statistics are functions that map the data histogram to some output range $\mathcal{R}$. A *database query* is $q : \mathbb{Z}^{*|\chi|} \to \mathcal{R}$. We call $q(x)$ the *exact query answer*.

The *sensitivity* of a query measures the maximum amount by which the exact answer can change when D changes by the addition or removal of exactly one row. The ℓ_1 sensitivity for query q is defined as

[16] In statistics, χ is called the sample space—the list of legal records in the database. Events that are deemed impossible a priori, structural zeros, can be accommodated in this framework. The assumption that χ is finite is not restrictive since, in practice, continuous data have discrete, bounded representations.

[17] If x is the histogram representation of D, y is the histogram representation of D', and D' is constructed from D by deleting or adding exactly one row, then $\|x - y\|_1 = 1$.

Definition 1 (ℓ_1 Query Sensitivity)

$$\Delta q = \max_{x,y \in \mathbb{Z}^{*|\mathcal{X}|}, \|x-y\|_1 \leq 1} |q(x) - q(y)|.$$

Sensitivity is a worst-case measure of how much information a given query can reveal. It is important in our analysis of privacy.

Most official statistical publications can be represented by linear queries. A *linear query* is $q(x) = q^T x$ where $q \in [-1, 1]^{|\mathcal{X}|}$. A *counting query* is a special case in which $q \in \{0, 1\}^{|\mathcal{X}|}$. Any marginal total from the fully-saturated contingency table representation of the database can be represented by a linear query. The tables for legislative redistricting, for example, are among millions of marginal tables published from the decennial census.

The statistical agency wants to publish answers to a *query workload*, $Q(\cdot) = \{q_1(\cdot), \ldots, q_k(\cdot)\}$. The exact answer to the query workload on the histogram x is a set $Q(x) = \{q_1(x), \ldots, q_k(x)\}$. In the absence of privacy concerns, the statistical agency would publish $Q(x)$. When the workload queries are linear, we represent the workload as a $k \times |\mathcal{X}|$ matrix Q, which is the vertical concatenation of the k scalar-valued linear queries. In this case, with some abuse of notation, we say Qx is the exact answer to the query workload $Q(x)$.[18] We extend Definition 1 to this setting in Section III.C.1.

II.B The Data Publication Mechanism

As in computer science, we model the data publication mechanism as a stochastic function.[19]

Definition 2 (Data Publication Mechanism) Let $\mathcal{F}$ be the set of allowable query work-

[18] It will be obvious from the context whether we are discussing the query workload, $Q(\cdot)$ or its matrix representation.

[19] Deterministic mechanisms are implicitly included as a special case, i.e., with probability one. Only trivial deterministic mechanisms are differentially private—"publish nothing" or "publish a constant." The distortion added to the exact query answer through the publication mechanism should enhance privacy, but will also reduce the accuracy of the published statistics.

loads. A *data publication mechanism* is a random function $M : \mathbb{Z}^{*|\mathcal{X}|} \times \mathcal{F} \to \mathcal{R}$ whose inputs are a histogram $x \in \mathbb{Z}^{*|\mathcal{X}|}$ and a workload $Q \in \mathcal{F}$, and whose random output is an element of range $\mathcal{R}$. For $B \in \mathcal{B}$, where $\mathcal{B}$ are the measurable subsets of $\mathcal{R}$, the conditional probability is $\Pr[M(x, Q) \in B | x, Q]$, given x and Q, where the probabilities are only over the randomness induced by the mechanism.

II.B.1 Differential Privacy

Our definition of differential privacy follows Dwork et al. (2006) and Dwork and Roth (2014).

Definition 3 (ε-differential privacy) Data publication mechanism M satisfies ε-differential privacy if for all $\varepsilon > 0$, all $x, x' \in N_x$, all $Q \in \mathcal{F}$, and all $B \in \mathcal{B}$

$$\Pr[M(x, Q) \in B \,|x, Q] \leq e^{\varepsilon} \Pr[M(x', Q) \in B \,|x', Q],$$

where $N_x = \left\{(x, x') \; s.t. \; x, x' \in \mathbb{Z}^{*|\mathcal{X}|} \text{ and } \|x - x'\|_1 = 1\right\}$ is the set of all *adjacent histograms* of x, and as in Definition 2 the probabilities are taken only over the randomness in the mechanism.[20]

II.B.2 Accuracy and Empirical Loss

We define accuracy in terms of the squared ℓ_2 distance between the mechanism output and the exact answer $Q(x)$.

Definition 4 (Accuracy (I)) Given histogram $x \in \mathbb{Z}^{*|\mathcal{X}|}$ and query workload $Q \in \mathcal{F}$, the data publication mechanism $M(x, Q)$ has accuracy I if

$$\mathbb{E}\left[\|(M(x, Q) - Q(x)\|_2^2\right] = -I.$$

[20]Mechanisms satisfying Definition 3 have several important properties. One that we use heavily is closure under composition, which means that if mechanism M_1 is ε_1-differentially private and mechanism M_2 is ε_2-differentially private, then the combination $M_{1,2}$ is $\varepsilon_1 + \varepsilon_2$ differentially private. In our case, the composed mechanism is $Q_1, Q_2 \in \mathcal{F}$ and $M_{1,2} \equiv M(x, [Q_1, Q_2])$. For a proof in the general case see Dwork and Roth (2014, Chapter 3).

where $I \leq 0$ and the expectation is taken only over the randomness in $M(x, Q)$.

The notation $\|\cdot\|_2^2$ is the square of the ℓ_2 (Euclidean) distance. Accuracy is usually defined in terms of a positive expected loss, $\alpha = -I$, rather than in terms of the additive inverse. We use the additive inverse, I, so we can model accuracy as a "good" rather than a "bad" in what follows. We make this normalization for rhetorical convenience, and it is without mathematical consequence.[21]

III Differentially Private Publication as a Production Technology

We explicitly characterize the production possibilities facing a statistical agency that publishes data using a known ε-differentially private mechanism. Doing so allows the agency to make an explicit guarantee regarding protection against social privacy loss; that is, to make a verifiable claim about the value of the parameter controlling the non-rival public good parameterized by ε. We describe two illustrative cases in which analysis of the mechanism yields a known technical relationship between accuracy and privacy loss. Furthermore, the function relating privacy loss and accuracy may be closed, bounded, and convex. When all these properties hold, there exists a known marginal rate of transformation between privacy loss and accuracy. It follows that the level of privacy loss chosen by the statistical agency entails an associated marginal cost of increasing privacy measured in units of foregone statistical accuracy.[22]

[21] Readers familiar with the computer science literature may wonder why we model accuracy in terms of the expected loss rather than the worst-case accuracy used, for example, in Dwork and Roth (2014). Expected squared-error loss is more familiar to economists. Our framework is readily extended to the other loss measures that appear in the differential privacy literature.

[22] Our analysis is not meant to be a descriptive, or positive, account of how statistical agencies or data custodians actually behave. It is explicitly normative.

III.A The Production of Privacy and Statistical Accuracy

Data publication mechanisms entail some bound on privacy loss (ε) and a level of statistical accuracy (I). Any mechanism is associated with a pair (ε, I). Following standard producer theory, we refer to each pair as a *production activity*. Production activities usually represent vectors of inputs and outputs such that the inputs can be transformed into the outputs. Our insight is to think of the information in the database as akin to an endowment of potential privacy loss. Some of the privacy loss endowment must be expended by the data custodian to produce population statistics of any accuracy.

The *transformation set*, Y, contains all feasible production activities available to the statistical agency. We assume Y is closed. We also assume inactivity is possible, but obtaining non-trivial accuracy requires some privacy loss (no free lunch). Likewise, obtaining perfect accuracy ($I = 0$) requires infinite privacy loss ($\varepsilon = \infty$). Under these assumptions, we can represent Y with a *transformation function* $G(\varepsilon, I)$ such that $Y = \{(\varepsilon, I) \,|\varepsilon > 0, I < 0 \text{ s.t. } G(\varepsilon, I) \leq 0\}$. The *production frontier* is the set

$$PF = \{(\varepsilon, I) \,|\varepsilon > 0, I < 0 \text{ s.t. } G(\varepsilon, I) = 0\}. \tag{1}$$

Equation (1) yields an implicit functional relationship between ε and I. As a theoretical proposition, (1) provides guidance on the constructs at the heart of our analysis. Its direct implementation is problematic. Given the current state of knowledge, discussed explicitly in Section III.D, there is no general solution for $G(\varepsilon, I)$. The statistical agency must select the best available technology for the query workload of interest and implement the G that technology implies. The agency should be guided by knowledge of recent advances and known impossibility results, but it cannot yet rely on algorithms known to solve equation (1) for general query workloads.

Scarcity is a key feature of the economic theory of production, often expressed in the axiom of "no free lunch." As it turns out, there is no free lunch when it comes to data privacy. In separate contributions, Dwork and Nissim (2004), Dwork and Naor (2010), Gehrke, Lui and Pass (2011), and Kifer and Machanava-

jjhala (2011) all show that publishing useful statistical summaries requires an explicit loss of privacy. This result holds for any non-trivial definition of formal privacy including, but not restricted to, differential privacy.

III.B Example: Randomized Response

To build intuition, we offer a brief, but accessible, example of a differentially private data publication mechanism known as *randomized response* (Warner 1965). The statistic of interest is the population proportion of a sensitive binary characteristic—for example, whether the respondent voted for a particular candidate. Randomized response protects the individual against privacy loss even when his "yes" or "no" response can be attributed directly to him. It does so by creating uncertainty about whether the respondent answered the sensitive question or some non-sensitive question.

In a survey setting, the respondent draws a sealed envelope. In it, the respondent finds one of two yes/no questions. The interviewer records the response, but does not know, and cannot record, which question was answered.[23] The data analyst knows only that the envelope contained the sensitive question with probability ρ and an innocuous question with probability $1 - \rho$.

Randomized response guarantees privacy to everyone. Those who answer the sensitive question can be assured that no one, including the interviewer, knows their response with certainty. We can measure the amount of information leaked about the sensitive characteristic. Finally, privacy increases with the probability that the innocuous question is asked. However, this also increases the uncertainty about the distance between the published statistic and the population proportion of the sensitive characteristic.

To formalize randomized response in terms of the model of this paper, suppose

[23] The sensitive data can also be collected automatically by a web browser, which performs the randomization in the background before transmitting information to an analyst. This approach is formalized in a tool known as Randomized Aggregatable Privacy-Preserving Ordinal Response (RAPPOR), and used by Google's Chrome browser to detect security problems (Erlingsson, Pihur and Korolova 2014; Fanti, Pihur and Erlingsson 2015).

the population database consists of two vectors x and z. Each entry x_i is a binary indicator corresponding to the sensitive trait for i. Each entry z_i corresponds to the non-sensitive trait. The statistical agency publishes a vector d that is conformable with x by randomizing over whether it reports the sensitive or the non-sensitive characteristic. Let T_i be a Bernoulli random variable that determines which entry is reported. The published value $d_i = T_i x_i + (1 - T_i)z_i$. To complete the description, denote $\rho = Pr[T_i = 1]$ and $\mu = Pr[z_i = 1]$. To make privacy guarantees, we require $0 < \rho < 1$. We also assume that z_i is independent of x_i. For convenience, we assume $\mu = 0.5$.[24]

Since the randomization is independent for each entry i, we can restrict attention to the conditional probabilities

$$Pr[d_i = 1|x_i = 1] = \rho + 0.5(1 - \rho) \tag{2}$$

$$Pr[d_i = 1|x_i = 0] = 0.5(1 - \rho). \tag{3}$$

Differential privacy bounds the ratio of these two probabilities as well as the ratio associated with the event $d_i = 0$. Randomized response is ε-differentially private with

$$\varepsilon(\rho) = \log\left(\frac{1+\rho}{1-\rho}\right). \tag{4}$$

Semantic, or inferential, privacy concerns what can be learned about the sensitive characteristic conditional on what is published, $Pr[x_i = 1|d_i]$. By Bayes' rule, the bound in (4) also applies to posterior inferences about the sensitive characteristic (see Appendix B).

The goal of the analysis is to draw inferences about the population proportion of the sensitive characteristic. Define $\widehat{\beta} = \frac{1}{N}\sum_i d_i$, the empirical mean proportion of ones in the published responses, and $\pi = \frac{1}{N}\sum_i x_i$, the (unobserved) mean of the sensitive characteristic. Finally, define $\widehat{\pi}(\rho) = \frac{\widehat{\beta}-\mu(1-\rho)}{\rho}$, which is an unbiased estimator of π with variance $\mathrm{Var}[\widehat{\pi}(\rho)] = \frac{\mathrm{Var}[\widehat{\beta}]}{\rho^2}$. Therefore, accuracy is: $I(\rho) = -\mathrm{Var}[\widehat{\pi}(\rho)]$.

[24]The justification for, and implications of, this assumption are elaborated in Appendix B. This assumption is without consequence for our illustration.

In Appendix B we also show $\frac{dI}{d\varepsilon} > 0$ and $\frac{d^2I}{d\varepsilon^2} < 0$. Therefore, the technical relationship between privacy-loss (ε) and accuracy (I) is strictly increasing and concave.

III.C Example: Matrix Mechanism

For the rest of the paper, we consider a statistical agency that publishes statistics using the *matrix mechanism* introduced by Chao Li, Gerome Miklau, Michael Hay, Andrew McGregor and Vibhor Rastogi (2015). Unlike randomized response, which operates directly on the micro-data record, the matrix mechanism is a general class of data-independent mechanisms that protect privacy by adding noise to the exact query answers that is calibrated to the query workload sensitivity. The matrix mechanism is also under active development for use by the Census Bureau (McKenna et al. 2018; Kuo et al. 2018). Our analysis is generally valid for all differentially private mechanisms that yield a convex relationship between privacy loss and accuracy.[25]

III.C.1 The Matrix Mechanism

To introduce the matrix mechanism, we first describe the simpler Laplace mechanism operating on a query workload. Dwork et al. (2006) proved the single query version of the Laplace mechanism, which Li et al. (2015) generalized to a matrix workload. We state the matrix version with new notation: ΔQ in Theorem 1 is the generalization of ℓ_1-sensitivity for the query workload (defined formally in Appendix C).

[25] In previous versions of this paper, we have considered data-dependent mechanisms, including the Multiplicative Weights Exponential Mechanism (MWEM) introduced by Moritz Hardt, Katrina Ligett and Frank McSherry (2012) and the Private Multiplicative Weights (PMW) mechanism, due to Moritz Hardt and Guy N. Rothblum (2010). While these mechanisms also yield well-behaved relationships between privacy and accuracy, their dependence on the data means their accuracy guarantees can only be stated in terms of the worst-case absolute error across all queries, not the expected square-error accuracy measure we focus on in this paper. This reinforces our observation that the definition of accuracy is not without loss of generality.

Theorem 1 (Laplace Mechanism) For $\varepsilon > 0$, query workload Q, and histogram x, define data publication mechanism $\text{Lap}(x, Q) \equiv Qx + e$, where e is a conformable vector of iid samples drawn from the Laplace distribution with scale parameter $b = \frac{\Delta Q}{\varepsilon}$. $Lap(x, Q)$ is ε-differentially private.

For the proof, see Li et al. (2015, prop. 2). The amount of noise added by the Laplace mechanism increases with the workload query sensitivity ΔQ and decreases with ε.[26]

The matrix mechanism improves on the Laplace mechanism by finding a query strategy matrix A, conformable with Q, such that each query in Q can be expressed as a linear combination of queries in A. The idea is to find a strategy matrix A with lower sensitivity than Q, thereby allowing greater accuracy for any particular level of privacy loss.

Theorem 2 (Matrix Mechanism Implemented with Laplace Mechanism) For histogram x, query workload Q, query strategy A, and $\varepsilon > 0$, the matrix mechanism $M(x, Q)$ publishes $Qx + QA^{+}(\Delta A)e$, where e is a vector of iid Laplace random variables with mean zero and scale parameter $b = 1/\varepsilon$.

1. *Privacy*: The matrix mechanism is ε-differentially private.

2. *Accuracy*: The matrix mechanism has accuracy $I = -\text{Var}(e)(\Delta A)^2 \|QA^{+}\|_F^2$, where A^{+} is the Moore-Penrose inverse of A, ΔA is the generalization of ℓ_1-sensitivity for the query workload (see Appendix C). The notation $\|\cdot\|_F$ refers to the matrix Frobenius norm, which is the square root of the sum of the absolute squared value of all elements in the vector or matrix between the braces (Golub and Van Loan 1996, p. 55).

For the proof, see Li et al. (2015, prop. 7).[27]

[26] A Laplace random variable—called "double exponential" by statisticians—has density $(1/2b)\exp(-|x|/b)$ on the real line with $\mathbb{E}[e] = 0$ and $\text{Var}[e] = 2b^2$. It is more peaked and has fatter tails than the normal distribution (Dwork and Roth 2014, p. 31). Its discrete equivalent is called the geometric distribution, and the associated geometric mechanism is also differentially private (Ghosh, Roughgarden and Sundararajan 2012, p. 1674-5).

[27] The strategy matrix A is not hypothetical. Li et al. (2015, p. 768-78) provide examples with and

When the queries in A are answered using the Laplace mechanism, the implied marginal cost of increasing accuracy I in terms of foregone privacy protection ε—the marginal rate of transformation—is

$$MRT \equiv \frac{dI}{d\varepsilon} = -\frac{\partial G/\partial \varepsilon}{\partial G/\partial I} = \frac{4\,(\Delta A)^2\,\|QA^+\|_F^2}{\varepsilon^3}. \tag{5}$$

The marginal rate of transformation is positive because privacy loss is a public bad.[28]

III.D Efficiency in Production

We would like to know whether the matrix mechanism provides maximal accuracy for any choice of ε, or if it is possible to achieve greater accuracy. For mechanisms involving multiple linear queries, Hardt and Talwar (2010) established upper and lower bounds on the variance of the noise added to achieve privacy. Their lower bound result implies there is a maximal attainable level of accuracy for any mechanism providing ε-differential privacy. They also established the existence of a differentially private mechanism, the *K-norm mechanism*, that approximately achieves maximal accuracy.[29] In principle, our matrix mechanism could be operated with the K-norm mechanism instead of the Laplace mechanism. However, there is no closed-form solution for accuracy when using the K-norm mechanism,

without side constraints of algorithms that successfully choose A. Ryan McKenna, Gerome Miklau, Michael Hay and Ashwin Machanavajjhala (2018) demonstrate the feasibility of computing A for large query workloads like those found in decennial censuses by exploiting Kronecker products. Nevertheless, we acknowledge that for general, large, linear query workloads, the computation of a solution for A is an unsolved problem.

[28] The *MRT* is not hypothetical. The accuracy guarantee in Definition (4) is exact. The production function is exact in the two public goods. The requirements $dI/d\varepsilon > 0$ and $d^2I/d^2\varepsilon < 0$ can be verified by substituting $\text{Var}(e) = 2/\varepsilon^2$ and differentiating. The matrix mechanism can be operated with any data-independent mechanism, such as the geometric mechanism, replacing the Laplace mechanism in its definition.

[29] Their results were subsequently refined and extended by Nikolov, Talwar and Zhang (2013) and Bhaskara et al. (2012).

and it is computationally burdensome.[30]

IV Preferences for Privacy and Social Choice

In this section we formally relate the definition of differential privacy to a semantic interpretation of privacy as the protection of individual secrets. We then develop a basic model of preferences for privacy and accuracy. We also derive the formal statement that the optimal levels of privacy protection and accuracy are determined by setting the marginal rate of transformation equal to the marginal willingness to accept privacy loss.

IV.A Differential Privacy as a Bound on Learning

Following Kifer and Machanavajjhala (2012), we describe necessary and sufficient conditions under which differential privacy implies a bound on what an attacker can learn about any person's sensitive data from published statistics. We call this the *semantic privacy* bound. We call the sensitive data items *secrets*. A secret pair consists of two mutually exclusive events—s_i and s_i'. The event s_i means the record for individual i was included in the database and has attribute χ_a. The event s_i' means the data for i was not included in the database.

Suppose the statistical agency publishes statistics using a ε-differentially private mechanism. Semantic privacy is the change in the odds of s_i versus s_i' for an attacker before and after the statistics are published. Inference requires a model of the random process that generates the database, $Pr[D \mid \theta]$. It is characterized by a parameter θ, and that the sampling probabilities for all individuals are independent of one another.[31] Applying Bayes' law to Definition 3 and using the data

[30] Computer scientists acknowledge the practical need to trade off computational costs against privacy and accuracy (Vadhan 2017, p. 50). We use the Laplace mechanism instead of the K-norm mechanism among data independent mechanisms for this reason.

[31] See Appendix D for details.

generating process in Appendix equation (D-13), differential privacy implies

$$e^{-\varepsilon} \leq \frac{Pr[s_i \,|B, \theta]}{Pr[s_i' \,|B, \theta]} \Bigg/ \frac{Pr[s_i \,|\theta]}{Pr[s_i' \,|\theta]} \leq e^{\varepsilon}. \tag{6}$$

For the proof, see Theorem 6.1 in Kifer and Machanavajjhala (2012).[32]

Equation (6) says that when the agency publishes using an ε-differentially private mechanism, the Bayes factor associated with any secret pair is bounded by the same ε. The semantics depend on the data generating process, and the conditioning on θ is nontrivial. As Kifer and Machanvajjhala show, the data generating process in equation (D-13) is the only one for which mechanisms satisfying differential privacy generate semantics satisfying equation (6).

Being able to move between the bounds on the mechanism and the bounds on the privacy semantics is critical to applying differential privacy to the social choice problem in this paper. When we derive technology sets, these bounds constrain the feasible pairs (ε, I); hence, the ε bounds on the mechanism implied by the definition of differential privacy (Definition 3) are important. When we model preferences for (ε, I), individuals care about the mechanism's ability to protect secrets learned from the published statistics; hence, the ε semantic bounds in equation (6) matter.

[32] Privacy semantics were defined as ε-indistinguishability, by Dwork (2006), one of the original differential privacy papers. This inference system is not Bayesian learning, and for our case, generates a semantic bound that is also $[-\varepsilon, \varepsilon]$ even though it is defined over arbitrary data generating processes. Kasiviswanathan and Smith (2014) also generate a semantic bound that can be computed from the differential privacy bound without assumptions on the data generating process. Wasserman and Zhou (2010) have given an interpretation of differential privacy as bounding the power of a statistical test of the null hypothesis about the value of a secret. Nissim et al. (2018, Section 4.3) provides an extremely lucid non-technical description of privacy semantics in terms of posterior-to-posterior comparisons.

IV.B Modeling Preferences over Privacy and Accuracy

We assume the statistical agency chooses ε and I to maximize a utilitarian social welfare function

$$SWF(\varepsilon, I) = \sum_i v_i(\varepsilon, I) \tag{7}$$

subject to the restriction that it operates along the production frontier (1). The functions $v_i(\varepsilon, I)$ measure indirect utility for each person i. Indirect utility depends on the levels of ε and I. Our formulation allows for arbitrary heterogeneity in preferences for both privacy loss and statistical accuracy. In doing so, we allow for the standard case in which one group cares only about privacy, while another group cares only about accuracy. Following Nissim, Orlandi and Smorodinsky (2012), we assume utility is additively separable into information utility and data utility: $v_i(\varepsilon, I) = v_i^{Info}(\varepsilon) + v_i^{Data}(I)$.[33]

IV.B.1 Information Utility

First, we specify individual information utility—preferences for privacy. Our approach is motivated by Ghosh and Roth (2015). Suppose that Ω is the set of future events, or states of the world, over which an individual has preferences. These are states of the world that may become more likely if confidential information in the database is disclosed. This might correspond to identity theft, the threat of being persecuted for a particular trait, or denial of a health insurance claim.

Let individual i's utility from event $\omega \in \Omega$ be $u_i(\omega)$. The individual's data may be used in an ε-differentially private mechanism $M(x, Q)$ with output drawn from range $\mathcal{R}$ according to the distribution induced by the mechanism. Finally, let $z(M(x, Q))$ be an arbitrary function that maps the published outcome of M onto a probability distribution over events in Ω. As discussed in Section I.A, differential privacy is invariant under post-processing. It follows that the transformation $z(M(x, Q))$ is also ε-differentially private because z only uses outputs of M.

[33] There is no obvious consumption complementarity between privacy and accuracy in the setup we consider. This assumption could be violated if, say, accuracy indirectly affects utility by increasing wealth or by decreasing the prices of physical goods.

From the individual's perspective, what matters is the difference between what can be learned about their secrets when their information is included in the data, x, and when it is not. Let x' denote the neighboring histogram that excludes i's data. By differential privacy, i's expected utility satisfies $\mathbb{E}_{\omega|M(x,Q)}[u_i(\omega)] \leq e^{\varepsilon}\mathbb{E}_{\omega|M(x',Q)}[u_i(\omega)]$. The worst-case incremental utility loss from having their data included in the mechanism, is $(e^{\varepsilon}-1)\,v_i$ where v_i is the expected utility over future events when i's data are not included in the mechanism. This argument supports a model of preferences in which ε enters linearly.[34] Following Ghosh and Roth (2015), we assume $v_i^{Info}(\varepsilon) = -k_i\varepsilon$ where $k_i \geq 0$ to reflect that privacy loss measured by ε is a public "bad."[35]

IV.B.2 Data Utility

Next, we introduce a model that supports a reduced-form specification for data utility that is linear in accuracy: $v_i^{data}(I) = a_i + b_i I$. Our simple, but illustrative, model is a proof-of-concept. We hope to inspire further investigation of preferences for public statistics, since there is very little existing research on which to draw.[36]

We associate the data utility for any person with her expected utility of wealth given her knowledge of the publication system. To model this expectation, we assume wealth depends on the statistics published by the statistical agency, and that individuals are aware that error is introduced by the privacy protection system. More concretely, each person i gets utility from wealth according to a twice-

[34]When ε is small, $\varepsilon \approx e^{\varepsilon} - 1$. Ghosh and Roth (2015) consider a broader class of information utility models, of which the linear model as a special case

[35]Nissim, Orlandi and Smorodinsky (2012) observe that the framework in Ghosh and Roth (2015) is an upper bound and that expected utility loss may be lower. They show that knowing the upper bound is sufficient for certain problems in mechanism design. The presence of ε in the utility function could also reflect the consumer's recognition that the statistical agency applies the same value of ε to everyone. We would assume they have preferences over this property of the mechanism. In this case, the connection to expected harm from breaches of their secrets would not be as direct.

[36]Spencer (1985); Spencer and Moses (1990) and Spencer and Seeskin (2015) attempt to measure the social cost of inaccuracy of official statistics. The statistics literature generally defines data "utility" via a loss function. See, e.g., Trottini and Fienberg (2002).

differentiable and strictly concave function, $U_i(W_i)$. We let $W_i = \Pi_i^T M(x, Q)$, where $M(x, Q)$ is the vector of published population statistics and Π_i is a person-specific vector of weights. We require only that the entries of Π_i be finite. As described in Theorem 2, $M(x, Q) = Qx + Q(\Delta A)A^+e$ where A is a query strategy matrix with pseudo-inverse A^+ and e is a vector of *iid* random variables with $\mathbb{E}[e] = 0$ and whose distribution is independent of x, Q, and A.

The query set Q and the distribution of e are public knowledge—a key feature of differential privacy. Hence, uncertainty is with respect to the data histogram x and the realized disturbances e. Beliefs about the distribution of x may be arbitrary and person-specific. The expected utility for any person i is

$$\mathbb{E}[U_i(W_i)] = \mathbb{E}_x\left[\mathbb{E}_{e|x}\left[U_i\left(\Pi_i^T Qx + \Pi_i^T QA^+(\Delta A)e\right)|x\right]\right].$$

In Appendx E, we show this can be approximated as

$$\begin{aligned}\mathbb{E}[U_i(W_i)] \approx\ & \mathbb{E}_x\left[U_i(\Pi_i^T Qx)\right] && -I \cdot \left\{\tfrac{1}{2}\mathbb{E}_x\left[U_i''(\Pi_i^T Qx)\right]\frac{\left\|\Pi_i^T QA^+\right\|_F^2}{\|QA^+\|_F^2}\right\} \\ \equiv\ & a_i && +I \cdot b_i.\end{aligned} \tag{8}$$

We obtain this result by taking expectations of a second-order Taylor series approximation to i's utility around $e = 0$. Our derivations and the result that expected utility is decreasing with the variance of wealth are familiar from the literature on risk aversion (Eeckhoudt, Gollier and Schlesinger 2005, see ch. 1). The final expression highlights that from the planner's perspective, all that matters are person-specific weights associated with the utility of data accuracy. We proceed with a reduced-form model for data utility that is linear in accuracy: $v_i^{data}(I) = a_i + b_i I$.

IV.B.3 Equilibrium

Assuming the indirect utility functions are differentiable, the conditions that characterize the welfare-maximizing levels of ε and I subject to the feasibility con-

straint are

$$\frac{\frac{\partial G\left(\varepsilon^0, I^0\right)}{\partial \varepsilon}}{\frac{\partial G\left(\varepsilon^0, I^0\right)}{\partial I}} = \frac{\frac{\partial}{\partial \varepsilon} \sum_i v_i^{Info}\left(\varepsilon^0\right)}{\frac{\partial}{\partial I} \sum_i v_i^{Data}\left(I^0\right)} = \frac{\sum_i k_i}{\sum_i b_i} = \frac{\bar{k}}{\bar{b}}. \tag{9}$$

The left-hand side of equation (9) is the marginal rate of transformation based on the production frontier while the right-hand side is the marginal rate of substitution between privacy loss and accuracy. We also refer to this as the *willingness to accept* privacy loss measured in units of statistical accuracy.

To choose privacy and accuracy, an agency needs to know the marginal rate of transformation and the willingness to accept as shown in (9). It must solve for the efficient query strategy A.[37] Once it does, it knows the marginal rate of transformation at any point. The choice then depends on two unknown quantities: the average preference for privacy in the population ($\bar{k}$) and the average preference for accuracy ($\bar{b}$). In Section V, we calibrate these quantities in an analysis that illustrates the strengths and drawbacks of our approach.

IV.C Accuracy and Privacy Protection Are Public Goods

In our model, ε and I are public goods. Once the statistical agency sets these parameters, all individuals enjoy the same statistics and privacy protection. However, our framework allows each person to have different preferences for privacy loss and accuracy.[38] It is natural to treat accuracy of official statistical publications as being both non-rival and non-excludable in consumption (Acquisti, Taylor and Wagman 2016). Hsu et al. (2014) and Ghosh and Roth (2015) model the provision of statistical summaries for use by a private analyst. Neither paper acknowledges the public-good nature of either the published statistics or the privacy protection afforded by the database custodian.

[37] The choice of query strategy A depends on the query workload, Q, the statistics of interest from the data collection, and the differentially private mechanism that will be used to answer A. See Li et al. (2015) for details.

[38] If statistical accuracy is an input to production, consumer utility also depends on accuracy indirectly through prices. Note, too, that our approach is distinct from much of the economics of privacy research that considers how companies elicit information about consumers' idiosyncratic preferences and thereby engage in price discrimination (Odlyzko 2004).

That privacy protection is a public good was foreshadowed by Dwork (2008, p. 3) when she wrote: "[t]he parameter ε ... is public. The choice of ε is essentially a social question." This observation has not previously made its way into models of the market for privacy rights. All persons receive the same guarantee against privacy loss, ε. That is, all persons participate in a data collection and publication system wherein the privacy-loss parameter is set independent of any characteristics of the population actually measured. This is our interpretation of the "equal protection under the law" confidentiality-protection constraint that governs most national statistical agencies. In addition, the benefits from increased privacy protection for any individual in the population are automatically enjoyed by every other individual, whether that person's information is used or not—the privacy protection is therefore strictly non-rival.

The public scrutiny of government statistical agencies create strong incentives to emphasize privacy protection (National Academies of Sciences, Engineering, and Medicine 2017, Chapter 5). In practice, an agency's choice is governed by legal, economic, and political considerations. One might reasonably ask "Why should the privacy bound ε arrived at by a statistical agency for reasons related to policy or legislation be the quantification of privacy loss relevant to economic actors? In environments where privacy loss is largely determined by a much smaller population of uniques or of people particularly susceptible to re-identification, features of that subpopulation might be all that determines the privacy-accuracy production function."[39] This argument is flawed by its failure to acknowledge that in all statistical databases of which we are aware every single row is unique.[40]

When all data are unique, deciding how to publish meaningful statistics involves choices that compromise the privacy of some sub-populations more than others. Conventional SDL attempts to present summaries with granularity in one dimension (say detailed geography) and similar summaries in another dimension (say detailed racial categories) without having to account for the risk in the

[39] We are grateful to an internal reviewer at the Census Bureau for providing this argument.

[40] For instance, in databases of households, the household address is geocoded to its GPS coordinates. In databases of businesses, detailed geography and industry codes uniquely distinguish firms.

cross-classification (detailed geography by detailed race categories). The database reconstruction theorem exposes the vulnerability in that practice. If every record is really a population unique, then publication mechanisms that are closed under composition are necessary to keep track of the exposure from the cumulative set of published statistics. Such mechanisms require worst-case analysis. And this worst-case analysis means privacy protection is a non-rival public good.

V Example: Title 1 School Funding Allocations

Our first application is the allocation of federal funding to school districts under Title I. If statistics on the number of Title I-eligible students in each school district are published using a differentially private mechanism, then funding will be misallocated due to the error induced by the mechanism. However, publishing those statistics without protection compromises privacy. We show how this social choice problem could be managed in practice.

V.A Setting

Title I of the Elementary and Secondary Education Act of 1965 provides federal funding to help improve educational outcomes for disadvantaged students. Funds are appropriated by Congress and then allocated to school districts based on need. That need is determined, in part, using statistics published by the Census Bureau.

Sonnenberg (2016) describes how Title I allocations are determined. The Department of Education (DOE) allocates basic grants using a formula that depends on $A_\ell = E_\ell \times C_\ell$, where A_ℓ is the *authorization amount* for school district ℓ, E_ℓ is the *eligibility count*, and C_ℓ is the *adjusted State Per-Pupil Expenditure (SPPE)*. To keep the analysis focused on the core privacy-accuracy trade-off, we assume this formula determines total funding to district ℓ and that C_ℓ is known with certainty, but the DOE must use a published count of Title I-eligible students $\widehat{E_\ell}$ that may differ

from the true count due to privacy protection.[41]

The DOE is supposed to allocate $X = \sum_{\ell=1}^{L} E_\ell \times C_\ell$ dollars under Title I, but the actual allocation will be $\widehat{X} = \sum_{\ell=1}^{L} \widehat{E}_\ell \times C_\ell$. The policy challenge is to balance privacy loss among disadvantaged households against the misallocation of Title I funds.

V.B The Social Choice Problem

We assume the Census Bureau has data on the complete population of school age children that includes two fields: the school district and an indicator for whether the child counts toward Title I eligibility. It publishes privacy-protected counts of the total number of Title I-eligible children in each district using the matrix mechanism of Theorem 2. The mechanism returns $\widehat{E}_\ell = E_\ell + e_\ell$ where e_ℓ is Laplace noise. By Theorem 2, the published counts satisfy ε-differential privacy.[42] We also know the accuracy is:

$$I = -\mathbb{E}\left[\sum_{\ell=1}^{L}\left(\widehat{E}_\ell - E_\ell\right)^2\right] = -\frac{2L}{\varepsilon^2} \tag{10}$$

where L is the total number of school districts.[43]

Suppose policymakers' choices are guided by a variant of the social welfare

[41] To be very clear, this does not describe the actual data collection or statistical disclosure practices currently used by the Census Bureau. Our example also abstracts from other types of non-survey error. These considerations are important, as Manski (2015) correctly argues. Our analysis shows that errors from privacy protection can be meaningful, even in the absence of other concerns about data quality.

[42] Because school districts do not overlap, the query workload sensitivity for the entire country is the same as the sensitivity for a single district—namely 1. The guarantee of ε-differential privacy is the same for each district separately—a property called *parallel composition*. The privacy loss for a student in one school district is not affected by what is guaranteed to students in other districts. This does not mean that learning the number of Title I students in district A is uninformative about Title I status of students in district B. If the districts are neighboring, the statistical outcomes may be correlated, and differentially private mechanisms permit learning about this correlation.

[43] Adding Laplace noise can result in published counts that are non-integer and potentially negative. Edits to enforce range restrictions—either by the Census Bureau prior to publication, or by the Department of Education prior to use—have only minor consequences for this analysis. The postprocessing does not affect the privacy guarantee, but it can affect the accuracy. See Li and Miklau (2012).

function studied in Section IV. Specifically:

$$SWF = \phi \sum_i v_i^{Info}(\varepsilon) + (1-\phi) v^{Data}(I), \tag{11}$$

where the first summand on the right-hand-side reflects the linear aggregation of individual utility from privacy loss, and the second summand reflects the social cost of misallocating funds. The parameter ϕ, which is new, is the weight on privacy in total welfare.

As in Section IV.B.1, let $v_i^{Info}(\varepsilon) = -k_i\varepsilon$ reflect the incremental loss to utility for person i associated with having her data used in a publication with privacy guarantee ε. Regarding $v^{Data}(I)$, the social planner's preferences are linear-quadratic in the aggregate misallocation $W = (\widehat{X} - X) = \sum_{\ell=1}^{L} C_\ell \left[\widehat{E}_\ell - E_\ell\right]$ so that $v^{Data}(I) = I \sum_{\ell=1}^{L} \frac{C_\ell^2}{L}$.[44] Following (9), the social planner's willingness to accept privacy loss is

$$WTA \equiv \frac{dI}{d\varepsilon} = \eta \frac{\sum_{i=1}^{N} k_i}{\overline{C^2}} = \eta N \frac{\bar{k}}{\overline{C^2}}, \tag{12}$$

where $\overline{C^2} = \frac{\sum_{\ell=1}^{L} C_\ell^2}{L}$ is average squared SPPE across districts, $\bar{k} = \frac{\sum_{i=1}^{N} k_i}{N}$ represents the average disutility from privacy loss across students, and N is the number of students. The parameter $\eta = \phi/(1-\phi)$ measures the relative weight on privacy in the social welfare function.

V.C Solution and Calibration

To establish a benchmark measure of WTA, we draw on the Common Core of Data (CCD) 2014—2015 (National Center for Education Statistics 2014). There are around $L = 13,000$ public school districts and $N = 46$ million school-age children. Following Sonnenberg (2016), we calculate the adjusted SPPE for each district.

[44] Here, we assume the planner has a distaste for misallocation. We could instead adopt a more general model for data utility following the analysis in Section IV.B.2 that would associate the planner's preferences for accuracy with the expected utility of each district. Doing so would add another set of utility parameters to model and calibrate and distract from our primary goal of offering a simple, stylized illustration of the approach developed in this paper.

The average squared SPPE, $\overline{C^2}$, is approximately 20 million.[45]

Completing the calibration requires information on preferences for privacy. In the Ghosh and Roth (2015) model, the privacy preference k_i is a measure of the loss associated with states of the world that become more likely when data are published. We posit the cost of identity theft as a reasonable reference point. We therefore set $\bar{k} = \$1,400$ based on a Department of Justice estimate of the direct and indirect losses for victims of identity theft who lost at least one dollar (Harrell 2017).[46] Hence, $WTA = \eta \times 1400 \times 2.3 = \eta \times 3220$.

Setting $WTA = MRT$ and making all relevant substitutions, $\varepsilon = 2.52 \times \eta^{-\frac{1}{3}}$. We report the cost of misallocation as the root mean squared error (RMSE) in expected allocation across districts.[47] The RMSE is measured in dollars per school district, so is comparable to other district-level expenditures.

We compare some benchmark models that place different relative weight on allocative efficiency. If privacy and accuracy are valued symmetrically, so $\eta = 1$, the optimal level of privacy loss is $\varepsilon^* = 2.52$, and the RMSE in allocations across districts is approximately \$2,509. This reflects a misallocation cost of about 70 cents per student. One might favor greater allocative efficiency, since the cost of misallocation affects everyone in the population. We might set $\eta = \frac{N}{POP-N} \approx 0.15$, where POP is the total U.S. population. Then $\varepsilon^{**} = 4.74$ and allocative inefficiency is just \$1,334, or approximately 38 cents per student.

Privacy advocates typically recommend values for ε that are less than 1 and much closer to zero (Dwork and Roth 2014). If we target $\varepsilon^0 = 0.1$, the RMSE in allocations across districts is approximately \$63,000. At roughly 18 dollars per student, the same amount would cover the cost of lunch for seven days.[48] The

[45] Using the CCD, we deduct federal outlays from total state education expenditures and divide by average fall daily attendance to get unadjusted SPPE. These are scaled down and truncated to get the adjusted SPPE according to the process described by Sonnenberg (2016). We then match each district to the adjusted SPPE of its home state.

[46] Technically, the estimated cost of identity theft reported by Harrell (2017) is \$1,343. In keeping with our goal of offering a simple, stylized analysis, we round up to the nearest hundred dollars.

[47] $RMSE = \sqrt{\mathbb{E}\left[L^{-1}\sum_{\ell=1}^{L} C_\ell^2\left(\widehat{E}_\ell - E_\ell\right)^2\right]} = \sqrt{-L^{-1}\overline{C^2}I}$.

[48] Based on a \$2.28 average cost per school lunch (U.S. Department of Agriculture, Food and

implied η is around 12,000.

VI Application to Current Statistical Programs

Our analysis focuses on how statistical agencies can optimally choose publication strategies that balance privacy and accuracy in a disciplined manner. However, most statistical agencies are not yet using formal privacy protection systems. In this section, we describe how the tools and concepts developed in this paper may be brought to bear on several real-world problems confronting the U.S. statistical system.

VI.A Legislative Redistricting

The Census Bureau is required to provide geographically detailed counts of the population to assist states in drawing legislative districts in compliance with the Equal Protection Clause of the 14^{th} Amendment of the U.S. Constitution and provisions of the 1965 Voting Rights Act. These PL94-171 redistricting statistics provide block-level population counts including data on race and ethnicity as mandated by the Office of Management and Budget (1997). The Census Bureau applies SDL to these tabulations; however, the procedures used the 2000 and 2010 Census, and those proposed for the 2018 End-to-End Census Test, do not protect the counts of total or voting-age population at any level of geography, including the block.[49] In implementing the amendments to the Census Act, the Census Bureau has acted as if the social welfare function put all the weight on accuracy when drawing fresh legislative districts. However, it has given weight to privacy protection in publishing the statistics used to enforce the Voting Rights Act.

An attacker with the information set posited in equation (6) can always correctly determine the census block of the person missing from that information set. Secret pairs like "Bob lives on block 1" versus "Bob is not in the database"

Nutrition Research Service, Office of Research, Nutrition and Analysis 2008).

[49]See Appendix A for the legislative and statutory background.

cannot be protected whereas secret pairs like "Bob is white (in the data)" versus "Bob left the race questions blank (in the data)" can. The absence of any provable privacy guarantee on the block-level geocode means it is impossible to provide any meaningful privacy protection on the geolocation in any public-use tables or micro-data published from the decennial census. The block geocodes are "blatantly non-private" (Dinur and Nissim 2003, p. 204). Hence, releasing block-level population counts without privacy protection still affects the Census Bureau's ability to protect other variables. The definition of neighboring databases must be changed. Only those matching the previously-published outputs are feasible.[50] By publishing the PL94-171 tables with exact population counts in all geographies, the Census Bureau has constrained its ability to protect confidentiality—it always exposes the block-level geocode from the confidential data.

VI.B Economic Censuses and National Accounts

We have assumed the set of statistics to publish (the query workload) is pre-specified. Producing tables with greater granularity reduces the accuracy of any specific table, holding privacy loss constant. How to choose the query workload is another open question.

Every five years the Census Bureau conducts an economic census of business establishments. A primary purpose is to provide statistical summaries against which the Bureau of Economic Analysis (BEA) benchmarks the national accounts. The BEA benchmarks take as inputs the same detailed geographic, industrial, and product summaries that the Census Bureau releases to the general public. These publications are prepared using traditional SDL methods. Regardless of the method, however, considerably more privacy protection is applied to these publications than would be required if the BEA used the Census Bureau's confidential inputs, and then applied the privacy protections directly to the national account summaries before publication. Such a publication technology would have greater accuracy and better privacy protection—it would Pareto dominate the sta-

[50] The applicable variant is *bounded* differential privacy (Kifer and Machanavajjhala 2011).

tus quo. Although the Census Bureau and the BEA are distinct agencies, CIPSEA permits the exchange of confidential business data between these agencies. The U.S. could achieve at least some of the efficiency gains that countries with consolidated statistical agencies achieve by relying more on the authorized data sharing provisions of that law.

VI.C Tax Data and Tax Simulations

Formal privacy systems allow for other innovative approaches to data dissemination. Consider the publication of data on individual, audited tax returns that are essential inputs for simulating the effects of tax policies (Feenberg and Coutts 1993).[51] The validity of tax simulations depends on each input record reflecting the exact relationship between income and tax liability. The Statistics of Income (SOI) Division of the Internal Revenue Service (IRS) has regularly published a public-use micro-data file of tax returns protected by micro-aggregation in which the data from small sets of input records, usually three, were pooled together (Harris-Kojetin et al. 2005, p. 49).[52] The micro-aggregation breaks some features of the audit consistency of the records and smooths over important breakpoints in the tax schedule, making them very hard to use in tax simulations. To release data suitable for tax simulations using differentially private mechanisms would require very large ε, even though tax returns are highly sensitive.

A different type of formal privacy protection could address both the technology and social choice for the tax simulation problem. SOI could run tax simulations behind its own firewall, using confidential tax returns as the inputs. The outputs of the tax simulation could be released using a differentially private publication system. Each user of the tax simulation system would have its own privacy-loss parameter, which SOI could control for global accounting of the privacy loss

[51] The Congressional Budget Office, the Joint Committee on Taxation, the Office of Tax Analysis (Treasury), the Federal Reserve System, many research organizations, and statistical agencies run tax simulations.

[52] The Statistics of Income Division currently sells these micro-data files, through 2012, for $4,000 (IRS Statistics of Income 2018). For documentation see National Bureau of Economic Research (2017).

associated with simulations. Since there is no requirement to hide the parameters of the differentially private publication system, the expected accuracy of the simulation outputs would be publicly known. These accuracy measures could then be used to make valid inferences about the effects of changing the inputs of the simulation. Errors induced by privacy protection could be considered in the same framework as other errors in the tax simulation. This solution to the social choice problem replaces an inefficient technology, traditional SDL, with an efficient one, formal privacy. It permits SOI to do the global accounting necessary to verify that its chosen weight on privacy protection relative to accuracy is properly implemented.[53]

VI.D General Purpose Public-use Micro-data

The Census Bureau published the first large-scale machine-readable public-use micro-data sample (PUMS) from the 1960 Census, selecting a one-percent sample of the records on the long form (Ruggles et al. 2011). The goal of the PUMS is to permit statistical analyses not supported by the published tabular statistics. Generating formally private microdata is a daunting challenge. A better strategy may be to develop new privacy-preserving approaches to problems that have historically been solved by PUMS.

One approach is an online system that interactively answers requests for tabular and statistical models. All such requests can be thought of as queries. When an agency can specify the set of allowable queries in advance, it is possible to design a formally private publication mechanism that operates on the confidential microdata and returns answers with known accuracy. A formally private PUMS would be dynamic, like the differentially private query systems embedded in Google's Chrome Browser, Apple's iOS 11, and Microsoft's Windows 10.[54]

An interactive online system works for models whose structure the agency

[53]The system proposed here relies on a differentially private interactive query system, such as those described by Gupta, Roth and Ullman (2012).

[54]For the Chrome browser see Fanti, Pihur and Erlingsson (2015). For iOS 11 see Differential Privacy Team (2017). For Windows 10 see Ding, Kulkarni and Yekhanin (2017).

can anticipate in advance (for example, the class of all linear regression models). More complicated analyses can be conducted in restricted-access environments. The Census Bureau has even acknowledged this publicly (U.S. Census Bureau 2018). Restricted-access environments don't automatically protect privacy. The data custodian still needs to decide how much of the privacy-loss budget to reserve for such unstructured studies, and the engineers still need to build formally private data analysis systems.[55]

VII Directions for Research

This paper has developed a coherent and rigorous framework for guiding decisions about how to effectively publish population statistics while preserving privacy. To make our framework practical will require more sophisticated models of production possibilities as well as better models, and measures, of the demand for privacy and accuracy. We briefly consider several promising extensions.

VII.A Extensions of the Production Model

We have focused on publication of contingency tables and other population aggregates. Our model of accuracy assumes that we only care about learning the finite-population statistic. This is not the same as learning about super-population parameters of the process that generated the data, as noted in Section IV. Our approach can be extended to statistical learning (Wasserman and Zhou 2010; Duchi, Jordan and Wainwright 2013; Dwork and Rothblum 2016). However, a robust approach must also allow for ambiguity regarding the true data generating process.

Our analysis of the economic census in Section VI.B suggests that the set of statistics to publish should be endogenous. Doing so requires a production technology that allows for different analyses to have different accuracy guarantees under the same publication system. For example, one could endow different users

[55] Several such systems have been developed: see McSherry (2009), Chen et al. (2016*a*), and Harvard Data Privacy Lab (2018).

with a fixed privacy-loss budget and let them submit queries. The answers would be accurate in proportion to the ratio of the query sensitivity to the privacy-loss endowment. More broadly, decisions about data collection, the set of tables to publish, and the privacy-loss budget, should be determined simultaneously. Our paper is a necessary step on the way toward a comprehensive analysis of decisions about collection, processing, and dissemination of information by statistical agencies.

We assume the statistical agency can explicitly enumerate its feasible combinations of privacy loss and statistical accuracy. This works when the statistical agency uses differentially private mechanisms with known accuracy and privacy guarantees. In more realistic settings where data are also subject to complex editing constraints, determining the production function is challenging. This is an active area of research.

VII.B Extensions of the Model of Preferences

The idea that statistical accuracy and privacy protection are public goods is not controversial, but does not often appear in models of data provision. We need models of market provision of statistics when those summaries are public goods. Such a model might start by extending the framework posed by Ghosh and Roth (2015) along the lines of Spence (1975), noting that those who sell their data value privacy less than the marginal consumer whose preferences a monopolistic provider will internalize.[56]

As Nissim, Orlandi and Smorodinsky (2012) point out, differential privacy only bounds the worst-case harm from participation. As one approach toward developing a better model, Kifer and Machanavajjhala (2012) suggest building data security algorithms based directly on privacy semantics. Turning to preferences for accuracy, we have assumed a reduced-form relationship between public statistics and wealth. A deeper analysis would directly model the role of public statistics in improving decision-making, as in Spencer (1985). Alternatively,

[56] We are grateful to an anonymous referee for pointing out this extension.

public statistics could enter production as a form of public capital, with accuracy affecting consumption through the demand for labor or reduced goods prices.

While differential privacy guarantees do not change over time, our static model abstracts from the possibility that the costs of privacy loss and the benefits of accurate statistics might time-vary. The Census Bureau's practice of making the complete responses from each decennial census public after 72 years is an implicit acknowledgment that privacy preferences (or, equivalently, the costs of privacy loss) change over time, at least relative to the social benefit of access to the full data. Further study is needed to determine how these dynamic considerations should factor into our basic social choice framework.

Finally, our models of data collection implicitly assume truthful reporting. With declining rates of survey participation, understanding the connection between data publication and data collection is also important. New thinking about who creates and who buys data, proposed in Arrieta-Ibarra et al. (2018), can also inform models of data acquisition by statistical agencies. Nissim, Orlandi and Smorodinsky (2012), Xiao (2013), and Chen et al. (2016*b*) also study the problem of eliciting truthful responses in the presence of privacy concerns.

VII.C The Need for Better Measurement

We need to learn more about preferences for privacy and accuracy. There is a growing body of evidence from public opinion surveys on attitudes toward privacy (Childs et al. 2012; Childs 2014; Childs, King and Fobia 2015). While these are instructive, it is far from clear how reported attitudes correspond to behavioral responses associated with changes in the risk of privacy loss. Some experiments have informed the value people attach to providing private data for commercial use (Acquisti, John and Loewenstein 2013). More information is needed on the price people attach to privacy loss, particularly as regards the inferential disclosures considered in this paper. With few exceptions (Spencer 1985; Mulry and Spencer 1993; Spencer and Seeskin 2015), there is virtually no evidence on the social value of public statistics, let alone the value of improving their accuracy.

VII.D Conclusion

Formal privacy models facilitate an interpretation of privacy protection as a commodity over which individuals have preferences. When statistical accuracy and privacy are public goods, as is the case for the publication of official statistics, their optimal levels are a social choice. This social choice can, and should, be guided by the principle of equating marginal social costs with marginal social benefits. In developing these ideas, we made many simplifying assumptions. We hope these can be excused as a feature of combining insights from three different disciplines to bear on a question of substantial scientific and public interest. We also hope our paper motivates researchers from economics, demography, computer science, statistics, and related disciplines to take up the Information Age challenge of designing publication systems that support accurate science but do not require infinite privacy loss.

References

13 U.S. Code. 1954. "USC: Title 13 - Census Act."

44 U.S. Code. 2002. "Confidential Information Protection and Statistical Efficency Act." Pub. L. 107-347, title V, Dec. 17, 2002, 116 Stat. 2962 (44 U.S.C. 3501 note).

Abowd, John M. 2016. "Why Statistical Agencies Need to Take Privacy-loss Budgets Seriously, and What It Means When They Do." *The 13th Biennial Federal Committee on Statistical Methodology (FCSM) Policy Conference*. https://digitalcommons.ilr.cornell.edu/ldi/32/.

Abowd, John M. 2017. "How Will Statistical Agencies Operate When All Data Are Private?" *Journal of Privacy and Confidentiality*, 7(3).

Abowd, John M., and Ian M. Schmutte. 2015. "Economic analysis and statistical disclosure limitation." *Brookings Papers on Economic Activity*, 221–267. Spring.

Acquisti, Alessandro, Curtis Taylor, and Liad Wagman. 2016. "The Economics of Privacy." *Journal of Economic Literature*, 54(2): 442–492.

Acquisti, Alessandro, Leslie K. John, and George Loewenstein. 2013. "What Is Privacy Worth?" *Journal of Legal Studies*, 42(2): 249–274.

Arrieta-Ibarra, Imanol, Leonard Goff, Diego Jimnez-Hernndez, Jaron Lanier, and E. Glen Weyl. 2018. "Should We Treat Data as Labor? Moving beyond "Free"." *AEA Papers and Proceedings*, 108: 38–42.

Bhaskara, Aditya, Daniel Dadush, Ravishankar Krishnaswamy, and Kunal Talwar. 2012. "Unconditional Differentially Private Mechanisms for Linear Queries." *STOC '12*, 1269–1284. New York, NY, USA:ACM.

Brenner, Hai, and Kobbi Nissim. 2014. "Impossibility of Differentially Private Universally Optimal Mechanisms." *SIAM Journal on Computing*, 43(5): 1513–1540.

Buckler, Warren. 1988. "Commentary: Continuous Work History Sample." *Social Security Bulletin*, 51(4): 12, 56.

Chen, Yan, Ashwin Machanavajjhala, Jerome P. Reiter, and Andres F. Barientos. 2016*a*. "Differentially Private Regression Diagnostics." *2016 IEEE International Conference on Data Mining*, 81–90.

Chen, Yiling, Stephen Chong, Ian A. Kash, Tal Moran, and Salil Vadhan. 2016*b*. "Truthful Mechanisms for Agents That Value Privacy." *ACM Trans. Econ. Comput.*, 4(3): 13:1–13:30.

Childs, Jennifer Hunter. 2014. "Understanding Trust in Official Statistics in the United States." Presentation at the 67th annual WAPOR conference in Nice, France in 2014. https://wapor.org/wp-content/uploads/WAPOR_Final_Program.pdf.

Childs, Jennifer Hunter, Ryan King, and Aleia Fobia. 2015. "Confidence in U.S. federal statistical agencies." *Survey Practice*, 8(5).

Childs, Jennifer Hunter, Stephanie Willson, Shelly Wilkie Martinez, Laura Rasmussen, and Monica Wroblewski. 2012. "Development of the Federal Statistical System Public Opinion Survey." http://www.aapor.org/AAPOR_Main/media/AnnualMeetingProceedings/2012/04_Childs-A6.pdf.

Commission on Evidence-Based Policymaking. 2017. "The Promise of Evidence-Based Policymaking: Report of the Commission on Evidence-Based Policymaking." Government Printing Office.

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

Data Stewardship Executive Policy Committee. 2014. "DS-22 Data Breach Policy Addendum." *https://www2.census.gov/foia/ds_policies/ds022.pdf*, Accessed on March 21, 2018.

Differential Privacy Team. 2017. "Learning with Privacy at Scale." *Apple Machine Learning Journal*, 1(8).

Ding, Bolin, Janardhan Kulkarni, and Sergey Yekhanin. 2017. "Collecting Telemetry Data Privately." *Advances in Neural Information Processing Systems 30*.

Dinur, Irit, and Kobbi Nissim. 2003. "Revealing information while preserving privacy." *Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, 202–210.

Donovan, Shaun. 2017. "Memorandum M-17-12 Preparing for and Responding to a Breach of Personally Identifiable Information." *https://obamawhitehouse.archives.gov/sites/default/files/omb/memoranda/2017/m-17-12_0.pdf*, Accessed on March 21, 2018.

Duchi, John C., Michael I. Jordan, and Martin J. Wainwright. 2013. "Local Privacy and Statistical Minimax Rates." *FOCS '13*, 429–438. Washington, DC, USA:IEEE Computer Society.

Duncan, George T., Mark Elliot, and Juan-José Salazar-González. 2011. *Statistical confidentiality principles and practice. Statistics for Social and Behavioral Sciences*, Springer New York.

Dwork, Cynthia. 2006. "Differential privacy." *Proceedings of the International Colloquium on Automata, Languages and Programming (ICALP)*, 1–12.

Dwork, Cynthia. 2008. "Differential privacy: a survey of results." *Theory and Applications of Models of Computation*, 1–19.

Dwork, Cynthia, Adam Smith, Thomas Steinke, Jonathan Ullman, and Salil Vadhan. 2015. "Robust traceability from trace amounts." 650–669. ACM Digital Library.

Dwork, Cynthia, and Aaron Roth. 2014. *The Algorithmic Foundations of Differential Privacy.* now publishers, Inc. Also published as "Foundations and Trends in Theoretical Computer Science" Vol. 9, Nos. 3–4 (2014) 211-407.

Dwork, Cynthia, and Guy N. Rothblum. 2016. "Concentrated differential privacy." *CoRR*, abs/1603.01887.

Dwork, Cynthia, and Kobbi Nissim. 2004. "Privacy-preserving datamining on vertically partitioned databases." *Proceedings of Advances in Cryptology (CRYPTO)*, 3152: 528–544.

Dwork, Cynthia, and Moni Naor. 2010. "On the difficulties of disclosure prevention in statistical databases or the case for differential privacy." *Journal of Privacy and Confidentiality*, 2(1): 93–107.

Dwork, Cynthia, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. "Calibrating Noise to Sensitivity in Private Data Analysis." *TCC'06*, 265–284. Berlin, Heidelberg:Springer-Verlag. DOI:10.1007/11681878_14.

Eeckhoudt, Louis, Christian Gollier, and Harris Schlesinger. 2005. *Economic and Financial Decisions Under Uncertainty.* Princeton University Press.

Erlingsson, Úlfar, Vasyl Pihur, and Aleksandra Korolova. 2014. "RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response." *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security - CCS '14*, 1054–1067.

Fanti, Giulia C., Vasyl Pihur, and Úlfar Erlingsson. 2015. "Building a RAPPOR with the unknown: privacy-preserving learning of associations and data dictionaries." *CoRR*, abs/1503.01214.

Feenberg, Daniel, and Elisabeth Coutts. 1993. "An Introduction to the TAXSIM Model." *Journal of Policy Analysis and Management*, 12(1): 189–194.

Garfinkel, Simson. 2015. "De-Identification of Personal Information." National Institute of Standards and Technology Internal Report 8053.

Gehrke, Johannes, Edward Lui, and Rafael Pass. 2011. "Towards privacy for social networks: A zero-knowledge based definition of privacy." 432–449, Springer.

Ghosh, Arpita, and Aaron Roth. 2015. "Selling privacy at auction." *Games and Economic Behavior*, 91: 334–346.

Ghosh, Arpita, Tim Roughgarden, and Mukund Sundararajan. 2012. "Universally Utility-maximizing Privacy Mechanisms." *SIAM Journal on Computing*, 41(6): 1673–1693.

Goldwasser, Shaft, and Silvio Micali. 1982. "Probabilistic encryption & how to play mental poker keeping secret all partial information." *STOC '82 Proceedings of the fourteenth annual ACM symposium on Theory of computing*, 365–377.

Golub, Gene H., and Charles F. Van Loan. 1996. *Matrix Computations, Third Edition.* The Johns Hopkins University Press.

Gupta, Anupam, Aaron Roth, and Jonathan Ullman. 2012. "Iterative constructions and private data release." *Lecture Notes in Computer Science (including sub-*

series Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7194 LNCS: 339–356.

Hardt, Moritz, and Guy N. Rothblum. 2010. "A Multiplicative Weights Mechanism for Privacy-Preserving Data Analysis." *2010 IEEE 51st Annual Symposium on 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 25.* , ed. F. Pereira, C.J.C. Burges, L. Bottou and K.Q. Weinberger, 2339–2347. Curran Associates, Inc.

Hardt, Mortiz, and Kunal Talwar. 2010. "On the Geometry of Differential Privacy." *STOC '10*, 705–714. ACM.

Harrell, Erika. 2017. "Victims of Identity Theft, 2014 (Revised November 13, 2017)." Department of Justice NCJ 248991.

Harris-Kojetin, Brian A., Wendy L. Alvey, Lynda Carlson, Steven B. Cohen, Steve H. Cohen, Lawrence H. Cox, Robert E. Fay, Ronald Fecso, Dennis Fixler, Gerald Gates, Barry Graubard, William Iwig, Arthur Kennickell, Nancy J. Kirkendall, Susan Schechter, Rolf R. Schmitt, Marilyn Seastrom, Monroe G. Sirken, Nancy L. Spruill, Clyde Tucker, Alan R. Tupek, G. David Williamson, and Robert Groves. 2005. "Statistical Policy Working Paper 22: Report on Statistical Disclosure Limitation Methodology." U.S. Federal Committee on Statistical Methodology Research Report.

Harvard Data Privacy Lab. 2018. "Harvard Data Privacy Lab Homepage." https://dataprivacylab.org/, Accessed: 2018-03-17.

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

Holan, Scott H., Daniell Toth, Marco A. R. Ferreira, and Alan F. Karr. 2010. "Bayesian Multiscale Multiple Imputation With Implications for Data Confidentiality." *Journal of the American Statistical Association*, 105(490): 564–577.

Homer, Nils, Szabolcs Szelinger, Margot Redman, David Duggan, Waibhav Tembe, Jill Muehling, John V. Pearson, Dietrich A. Stephan, Stanley F. Nelson, and David W. Craig. 2008. "Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays." *PLOS Genetics*, 4(8): 1–9.

Hsu, Justin, Marco Gaboardi, Andreas Haeberlen, Sanjeev Khanna, Arjun Narayan, Benjamin C. Pierce, and Aaron Roth. 2014. "Differential Privacy: An Economic Method for Choosing Epsilon." *2014 IEEE 27th Computer Security Foundations Symposium*, 398–410.

IRS Statistics of Income. 2018. "SOI Products and Services." http://www.irs.gov/statistics/soi-tax-stats-individual-public-use-microdata-files, Accessed on July 31, 2018.

Jones, Christa. 2017. "Nonconfidential Memorandum on Census Bureau Privacy Breaches." *Memorandum to file*, public document in replication archive 10.5281/zenodo.1208758.

Kasiviswanathan, Shiva P, and Adam Smith. 2014. "On the 'Semantics' of Differential Privacy: A Bayesian Formulation." *Journal of Privacy and Confidentiality*, 6(1): 1.

Kifer, Daniel, and Ashwin Machanavajjhala. 2011. "No free lunch in data privacy." *SIGMOD '11*, 193–204. New York, NY, USA:ACM Digital Library.

Kifer, Daniel, and Ashwin Machanavajjhala. 2012. "A rigorous and customizable framework for privacy." *Proceedings of the 31st symposium on Principles of Database Systems - PODS '12*, 77.

Kuo, Yu-Hsuan, Cho-Chun Chiu, Daniel Kifer, Michael Hay, and Ashwin Machanavajjhala. 2018. "Differentially Private Hierarchical Group Size Estimation." *CoRR*, abs/1804.00370.

Li, Chao, and Gerome Miklau. 2012. "An adaptive mechanism for accurate query answering under differential privacy." *Proceedings of the VLDB Endowment*, 5(6): 514–525.

Li, Chao, Gerome Miklau, Michael Hay, Andrew McGregor, and Vibhor Rastogi. 2015. "The matrix mechanism: optimizing linear counting queries under differential privacy." *The VLDB Journal*, 24(6): 757–781.

Mandel, B. J. 1950. "OASI Earnings Statistics and Their Uses." *Monthly Labor Review*, 70(4): 421–425.

Manski, Charles F. 2015. "Communicating Uncertainty in Official Economic Statistics: An Appraisal Fifty Years after Morgenstern." *Journal of Economic Literature*, 53(3): 631–53.

McKenna, Ryan, Gerome Miklau, Michael Hay, and Ashwin Machanavajjhala. 2018. "Optimizing error of high-dimensional statistical queries under differential privacy." *Proceedings of the VLDB Endowment*, 11(10).

McSherry, Frank. 2009. "Privacy Integrated Queries." *Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data (SIGMOD)*.

Mulry, Mary H., and Bruce D. Spencer. 1993. "Accuracy of the 1990 Census and Undercount Adjustments." *Journal of the American Statistical Association*, 88(423): 1080–1091.

Narayanan, Arvind, and Vitaly Shmatikov. 2008. "Robust De-anonymization of Large Sparse Datasets." *SP '08*, 111–125. Washington, DC, USA:IEEE Computer Society. DOI:10.1109/SP.2008.33.

National Academies of Sciences, Engineering, and Medicine. 2017. *Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy. Committee on National Statistics*, Washington, DC:National Academies Press.

National Bureau of Economic Research. 2017. "U.S. Individual Income Tax Public Use Sample Documentation." https://http://users.nber.org/~taxsim/gdb/, Accessed on July 31, 2018.

National Center for Education Statistics. 2014. "Common Core of Data." U.S. Department of Education [Computer file] v.1a.

National Institutes of Health. 2014. "NIH Genomic Data Sharing Policy." https://grants.nih.gov/grants/guide/notice-files/NOT-OD-14-124.html, Accessed: March 13, 2018.

Nikolov, Aleksandar, Kunal Talwar, and Li Zhang. 2013. "The Geometry of Differential Privacy: The Sparse and Approximate Cases." *STOC '13*, 351–360. ACM.

Nissim, Kobbi, Claudio Orlandi, and Rann Smorodinsky. 2012. "Privacy-aware mechanism design." *EC '12*, 774–789. New York, NY, USA:ACM.

Nissim, Kobbi, Thomas Steinke, Alexandra Wood, Micah Altman, Aaron Bembenek, Mark Bun, Marco Gaboardi, David R. O'Brien, and Salil Vadhan. 2018. "Differential Privacy: A Primer for a Non-Technical Audience." *Privacy Law Scholars Conference 2017*.

Odlyzko, Andrew. 2004. "Privacy, Economics, and Price Discrimination on the Internet." *Economics of Information Security*, , ed. L. Jean Camp and Stephen Lewis, 187–211. Boston, MA:Springer US.

Office of Management and Budget. 1997. "Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity." https://obamawhitehouse.archives.gov/omb/fedreg_1997standards, (Cited on March 22, 2018).

Perlman, Jacob. 1951. "The Continuous Work-History Sample: The First 12 Years." *Social Security Bulletin*, 14(4): 3–10.

Perlman, Jacob, and Benjamin Mandel. 1944. "The Continuous Work History Sample Under Old-Age and Survivors Insurance." *Social Security Bulletin*, 7(2): 12–22.

Ruggles, Steven, Matthew Schroeder, Natasha Rivers, J. Trent Alexander, and Todd K. Gardner. 2011. "Frozen Film and FOSDIC Forms: Restoring the 1960 U.S. Census of Population and Housing." *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 44(2): 69–78.

Samuelson, Paul A. 1954. "The pure theory of public expenditure." *Review of Economics and Statistics*, 37: 387–389.

Smith, Chreston M. 1989. "The Social Security Administration's Continuous Work History Sample." *Social Security Bulletin*, 52(10): 20–28.

Sonnenberg, William. 2016. "Allocating Grants for Title I." National Center for Education Statistics.

Spence, A. Michael. 1975. "Monopoly, Quality, and Regulation." *The Bell Journal of Economics*, 6(2): 417–429.

Spencer, Bruce D. 1985. "Optimal Data Quality." *Journal of the American Statistical Association*, 80(391): 564–573.

Spencer, Bruce D., and Lincoln E. Moses. 1990. "Needed Data Expenditure for an Ambiguous Decision Problem." *Journal of the American Statistical Association*, 85(412): 1099–1104.

Spencer, Bruce David, and Zachary H. Seeskin. 2015. "Effects of Census Accuracy on Apportionment of Congress and Allocations of Federal Funds." *JSM Proceedings, Government Statistics Section*, 3061–3075.

Topel, Robert H., and Michael P. Ward. 1992. "Job Mobility and the Careers of Young Men." *The Quarterly Journal of Economics*, 107(2): 439–479.

Trottini, Mario, and Stephen E. Fienberg. 2002. "Modelling User Uncertainty for Disclosure Risk and Data Utility." *International Journal of Uncertainty and Fuzziness in Knowledge-Based Systems*, 10(5): 511–527.

U.S. Census Bureau. 2002. "Census Confidentiality and Privacy 1790 to 2002." https://www.census.gov/prod/2003pubs/conmono2.pdf, (Cited on March 22, 2018).

U.S. Census Bureau. 2012. "2010 Census Summary File1–Technical Documentation." Department of Commerce, Economics and Statistics Administration.

U.S. Census Bureau. 2017. "Census Scientific Advisory Committee." U.S. Commerce Department. https://www.census.gov/about/cac/sac/meetings/2017-09-meeting.html (Cited on March 22, 2018).

U.S. Census Bureau. 2018. "Restricted-Access Microdata." https://www.census.gov/research/data/restricted_use_microdata.html, Accessed: 2018-03-17.

U.S. Department of Agriculture, Food and Nutrition Research Service, Office of Research, Nutrition and Analysis. 2008. "School Lunch and Breakfast Cost Study II, Final Report." Special Nutition Programs CN-08-MCII. by Susan Bartlett, Frederic Glantz, and Christopher Logan.

U.S. Department of Commerce. 2018. "U.S. Department of Commerce Announces Reinstatement of Citizenship Question to the 2020 Decennial Census." https://www.commerce.gov/news/press-releases/2018/03/us-department-commerce-announces-reinstatement-citizenship-question-20
Accessed: March 13, 2018.

U.S. Supreme Court. 1999. "DEPARTMENT OF COMMERCE v. UNITED STATES HOUSE (98-404) No. 98—404, 11 F. Supp. 2d 76, appeal dismissed; No. 98—564,

19 F. Supp. 2d 543, affirmed." https://www.law.cornell.edu/supct/html/98-404.ZO.html, (Cited on March 26, 2018).

U.S. Supreme Court. 2002. "UTAH V. EVANS (01-714) 536 U.S. 452 182 F. Supp. 2d 1165, affirmed." https://www.law.cornell.edu/supct/html/01-714.ZS.html, (Cited on March 22, 2018).

Vadhan, Salil. 2017. "The Complexity of Differential Privacy." *Tutorials on the Foundations of Cryptography: Dedicated to Oded Goldreich*, , ed. Yehuda Lindell, 347–450. Springer International Publishing.

Warner, Stanley L. 1965. "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias." *Journal of the American Statistical Association*, 60(309): 63–69.

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?" *ITCS '13*, 67–86. New York, NY, USA:ACM.

Yu, Fei, Stephen E. Fienberg, Aleksandra B. Slavkovic, and Caroline Uhler. 2014. "Scalable privacy-preserving data sharing methodology for genome-wide association studies." *Journal of Biomedical Informatics*, 50: 133 – 141. Special Issue on Informatics Methods in Medical Privacy.

An Economic Analysis of Privacy Protection and Statistical Accuracy as Social Choices

John M. Abowd and Ian M. Schmutte

Online Appendix

August 15, 2018

A Potential Database Reconstruction Attack against the 2010 Decennial Census

In Section I.B.3 we discuss the potential for a database reconstruction attack against the decennial census based on the large number of summary tables published from the confidential micro-data. Using the schema in the public documentation for PL94-171, Summary File 1, Summary File 2, and the Public-use Micro-data Sample, and summarizing from the published tables, there were at least 2.8 billion *linearly independent* statistics in PL94-171, 2.8 billion in the balance of SF1, 2.1 billion in SF2, and 31 million in the PUMS https://www.census.gov/prod/www/decennial.html (cited on March 17, 2018). For the 2010 Census, the national sample space at the person level has approximately 500,000 cells. The unrestricted sample space at the census block level has approximately $500{,}000 \times 10^7$ cells. It might seem there are orders of magnitude more unknowns than equations in the system used for reconstruction. However, traditional SDL does not protect sample zeros. Consequently, every zero in a block, tract, or county-level table rules out all record images in the sample space that could have populated that cell, dramatically reducing the number of unknowns in the relevant equation system.

The deliberate preservation of sample zeros can be inferred from the technical documentation: "Data swapping is a method of disclosure avoidance designed to protect confidentiality in tables of frequency data (the number or percentage of the population with certain characteristics). Data swapping is done by editing the source data or exchanging records for a sample of cases. A sample of households is selected and matched on a set of selected key variables with households in neighboring geographic areas (geographic areas with a small population) that have similar characteristics (same number of adults, same number of children, etc.). Because the swap often occurs within a geographic area with a small population, there is no effect on the marginal totals for the geographic area with a small population or for totals that include data from multiple geographic areas with small populations. Because of data swapping, users should not assume that tables with cells having a value of one or two reveal information about specific individuals" (U.S. Census Bureau 2012, p. 7-6).

B Randomized Response Details

A custodian collects data from a population of individuals, $i \in \{1, \ldots, N\}$. Each member of the population has a sensitive characteristic and an innocuous characteristic. The sensitive characteristic is $x_i = Y_i(1) \in \{0, 1\}$, with population proportion $\Pr[Y_i(1) = 1] = \pi$. This proportion, π, is the unknown population quantity of interest. The non-sensitive characteristic is $z_i = Y_i(0) \in \{0, 1\}$ with known population proportion $\Pr[Y_i(0) = 1] = \mu$. The custodian collects and publishes a mixture

$$d_i = T_i Y_i(1) + (1 - T_i) Y_i(0), \tag{B-1}$$

where T_i indicates whether the sensitive or the non-sensitive question was collected, with $\Pr[T_i = 1] = \rho$. The responses are independent of which information is collected: $(Y_i(1), Y_i(0)) \perp\!\!\!\perp T_i$. We also require that the non-sensitive item be independent of the sensitive item. This is not restrictive, since the innocuous question can literally be "flip a coin and report whether it came up heads," as in the original

application.

The indicator T_i is not observed. Any data analyst observes only the reported variable d_i. However, as in a randomized controlled trial, the probability of T_i, ρ, is known with certainty. Furthermore, the analyst also knows the probability of the non-sensitive response, μ.

Define $\widehat{\beta} = \frac{1}{N}\sum_i d_i$, the empirical mean proportion of responses of one. Independence of T_i implies $\mathrm{E}\left[\widehat{\beta}\right] = \pi\rho + \mu(1-\rho)$. It follows that $\widehat{\pi} = \frac{\widehat{\beta}-\mu(1-\rho)}{\rho}$ is an unbiased estimator of π with variance $\mathrm{Var}[\widehat{\pi}] = \mathrm{Var}[\widehat{\beta}]\rho^{-2}$.

B1 Privacy under Randomized Response

For given ε, differential privacy requires both $\Pr[d_i = 1|Y_i(1) = 1] \leq e^{\varepsilon}\Pr[d_i = 1|Y_i(1) = 0]$, and $\Pr[d_i = 0|Y_i(1) = 0] \leq e^{\varepsilon}\Pr[d_i = 0|Y_i(1) = 1]$. Together, these expressions bound the Bayes factor, which limits how much can learned about the sensitive characteristic upon observation of the collected response.

Making substitutions based on the data-generating model,

$$1 + \frac{\rho}{(1-\rho)\mu} \leq e^{\varepsilon} \tag{B-2}$$

and

$$1 + \frac{\rho}{(1-\rho)(1-\mu)} \leq e^{\varepsilon}. \tag{B-3}$$

For a given choice of μ, the differential privacy guaranteed by randomized response is the maximum of the values of the left-hand sides of equations (B-2) and (B-3). Hence, privacy loss is minimized when $\mu = \frac{1}{2}$. This is the case we will consider throughout the remaining discussion. We note doing so assumes that inferences about affirmative and negative responses are equally sensitive, which may not always be the case. The results of our analysis do not depend on this assumption. [57]

[57] These observations highlight another allocation problem: how to trade off protection of affirmative responses for the sensitive item $Y_i(1) = 1$ against protection of negative responses $Y_i(1) = 0$. What do we mean? If ρ is fixed, then increasing μ reduces the Bayes factor in (B-2) (increasing

For randomized response, the differential privacy guarantee as a function of ρ is:

$$\varepsilon(\rho) = \log\left(1 + \frac{2\rho}{(1-\rho)}\right) = \log\left(\frac{1+\rho}{1-\rho}\right), \tag{B-4}$$

which follows from setting $\mu = \frac{1}{2}$ in equations (B-2) and (B-3).

B2 Statistical Accuracy under Randomized Response

Expressed as a function of ρ, we denote the estimated share of the population with the sensitive characteristic

$$\widehat{\pi}(\rho) = \frac{\widehat{\beta}(\rho) - \mu(1-\rho)}{\rho} \tag{B-5}$$

where $\widehat{\beta}(\rho)$ is the population average response when the sensitive question is asked with probability ρ. Clearly,

$$\mathrm{E}[\widehat{\beta}(\rho)] = [\rho\pi + (1-\rho)\mu] \tag{B-6}$$

and

$$\mathrm{Var}[\widehat{\beta}(\rho)] = \frac{[\rho(\pi-\mu)+\mu](1-\rho(\pi-\mu)-\mu)}{N}. \tag{B-7}$$

privacy) and increases the Bayes factor in (B-3) (decreasing privacy). The underlying intuition is fairly simple. Suppose the sensitive question is "did you lie on your taxes last year?" Most tax evaders would prefer that their answer not be made public, but non-evaders are probably happy to let the world know they did not cheat on their taxes. In such a setting, with ρ fixed, we can maximize privacy for the tax evader by setting μ to 1. Recall μ is the probability of a positive response on the non-sensitive item ($Y_i(0) = 1$). If $\mu = 1$, then when the data report $d_i = 0, 1$ we know with certainty that $Y_i(1) = 0$ (i.e., i did not cheat on her taxes). In this special case, the mechanism provides no privacy against inference regarding non-evasion, but maximum attainable privacy (given the mechanism) against inference regarding evasion. This is the role the Bloom filter plays in the full RAPPOR implementation of randomized response (Erlingsson, Pihur and Korolova 2014). More generally, the choice of μ can be tuned to provide relatively more or less privacy against one inference or the other.

It follows that

$$\text{Var}[\widehat{\pi}(\rho)] = \frac{\text{Var}[\widehat{\beta}(\rho)]}{\rho^2} = \frac{[\rho(\pi-\mu)+\mu](1-\rho(\pi-\mu)-\mu)}{\rho^2 N}. \tag{B-8}$$

We can define data quality as:

$$I(\rho) = \text{Var}[\widehat{\pi}(1)] - \text{Var}[\widehat{\pi}(\rho)]. \tag{B-9}$$

This measures the deviation in the sampling variance for the predicted population parameter, π, relative to the case where there is no privacy protection ($\rho = 1$).

B3 The Accuracy Cost of Enhanced Privacy under Randomized Response

Equations (B-4) and (B-9) implicitly define a functional relationship between data privacy, parameterized by ε, and accuracy, parameterized as I. This function tells us the marginal cost borne by individuals in the database necessary to achieve an increase in accuracy of the published statistics. We can characterize the relationship between accuracy, I, and privacy loss, ε, analytically. First, we invert equation (B-4) to get ρ as a function of ε:

$$\rho(\varepsilon) = \frac{e^{\varepsilon}-1}{1+e^{\varepsilon}}. \tag{B-10}$$

Next, we differentiate I with respect to ε via the chain rule: $\frac{dI}{d\varepsilon} = I'(\rho(\varepsilon))\rho'(\varepsilon)$:

$$I'(\rho) = \frac{2\,\text{Var}[\widehat{\beta}(\rho)]}{\rho} - \frac{(\pi-\frac{1}{2})(1-2\pi)}{N^2\rho}. \tag{B-11}$$

and

$$\rho'(\varepsilon) = \frac{2e^{\varepsilon}}{(1+e^{\varepsilon})^2} = \frac{1}{1+cosh(\varepsilon)}. \tag{B-12}$$

Both derivatives are positive, so it follows that $\frac{dI}{d\varepsilon} > 0$. A similar derivation shows that $\frac{d^2I}{d\varepsilon^2} < 0$. Increasing published accuracy requires an increase in privacy loss

at a rate given by $\frac{dI}{d\varepsilon} > 0$. Furthermore, achieving a given increment in accuracy requires increasingly large privacy losses.

C Details of the Matrix Mechanism

For a single query, we defined the ℓ_1 sensitivity in Definition 1. The results in Theorems 1 and 2 are defined in terms of the sensitivity of a workload of linear queries, which we denote ΔQ. Following Li et al. (2015),

Theorem 3 (ℓ_1 Query Matrix Sensitivity) Define the ℓ_1 sensitivity of Q by

$$\Delta Q = \max_{x,y \in \mathbb{Z}^{*|k|}, \|x-y\|_1 \leq 1} \|Qx - Qy\|_1 .$$

This is equivalent to

$$\Delta Q = \max_k \|q_k\|_1 ,$$

where q_k are the columns of Q.

For the proof, see Li et al. (2015, prop. 4).

D Details of Privacy Semantics

We provide technical definitions associated with the derivations in Kifer and Machanavajjhala (2012) described in Section IV.A.

Assume a latent population of individuals $h_i \in \mathcal{H}$ of size N^*. The confidential database, D, is a random selection of $N < N^*$ individuals, drawn independently from $\mathcal{H}$. In this context N is a random variable, too. The database also records characteristics of each individual, which are drawn from the data domain χ. Denote the record of individual i as r_i. The event "the record r_i is included in database D" has probability π_i. Denote the conditional probability of the event "the record $r_i = \chi_a \in \chi$" given that r_i is in D as $f_i(r_i)$. Then, the data generating process is parameterized by $\theta = \{\pi_1, ..., \pi_N, f_1, ..., f_N\}$. The probability of database D, given θ,

is

$$Pr\left[D \mid \theta\right] = \prod_{h_i \in D} f_i(r_i)\pi_i \prod_{h_j \notin D}(1-\pi_j). \tag{D-13}$$

The complete set of paired hypotheses that differential privacy protects is

$$\mathcal{S}_{pairs} = \{(s_i, s_i') : h_i \in \mathcal{H}, \chi_a \in \chi\}, \tag{D-14}$$

where s and s' are defined in Section IV.A. By construction $\mathcal{S}_{pairs}$ contains every pair of hypotheses that constitute a potential disclosure; that is, whether any individual h_i from the latent population is in or out of the database D and, if in D, has record r_i.

E Derivation of the Data Utility Model

Recall that the matrix mechanism publishes a vector of answers, $M(x, Q)$ to the known set of queries, Q given an underlying data histogram x. The matrix mechanism is implemented by using a data independent mechanism to answer a set of queries represented by the query strategy matrix, A with sensitivity ΔA and pseudo-inverse A^+. Following Theorem 2, $M(x, Q) = Qx + Q(\Delta A)A^+e$ where e is a vector of *iid* random variables with $\mathbb{E}\left[e\right] = 0$ and whose distribution is independent of x, Q, and A. In what follows, we use the notation σ_e^2 to denote the common (scalar) variance of the elements of e. For example, when e is a vector of Laplace random variables with scale ε^{-1}, we know that $\sigma_e^2 = 2\varepsilon^{-2}$. Note that the variance of the vector e is $\mathbb{E}\left[ee^T\right] = \sigma_e^2\mathbb{I}$ where $\mathbb{I}$ is the identity matrix conformable with e.

Let $W_i = \Pi_i^T M(x, Q)$ be a person-specific linear function by which published statistics are transformed into wealth (or consumption). Individuals have utility of wealth given by a twice-differentiable and strictly concave function, $U_i(W_i)$. The total realized ex post wealth for i is $W_i = \Pi_i^T Qx + \Pi_i^T QA^+(\Delta A)e$. We assume i knows Q and the details of the mechanism M. Uncertainty is over x and e.

For notational convenience, we define a function $w_i(e; x) = \Pi_i^T Qx + \Pi_i^T QA^+(\Delta A)e$. Conditional on x, the expected utility of i from receiving the mechanism output

is $\mathbb{E}_{e|x}\left[U_i\left(w_i(e;x)\right)|x\right]$. We approximate this by taking expectations of a second-order Taylor Series expansion of $h_i(e;x) = U_i\left(w_i(e;x)\right)$ with respect to e evaluated at $e_0 = 0$.

Let $\nabla h_i(e_0;x)$ denote the gradient of h with respect to e and let $H_i(e_0;x)$ denote the Hessian. The second-order Taylor series expansion of $h_i(e;x)$ evaluated at e_0 is

$$h_i(e;x) \approx h_i(e_0;x) + (e-e_0)^T \nabla h_i(e_0;x) + \frac{1}{2!}(e-e_0)^T H_i(e_0;x)(e-e_0). \tag{E-15}$$

The gradient of h is

$$\nabla h_i(e_0;x) = U_i'(w_i(e_0;x))\Delta A\left(\Pi_i^T Q A^+\right)^T. \tag{E-16}$$

The Hessian is

$$H_i(e_0;x) = U_i''(w_i(e_0;x))(\Delta A)^2\left(\Pi_i^T Q A^+\right)^T\left(\Pi_i^T Q A^+\right). \tag{E-17}$$

Note that we have used the chain rule in both derivations. We now evaluate the right hand side of equation (E-15) at $e_0 = 0$. Defining new notation, let $w_{i0}^x = w_i(0;x) = \Pi_i^T Q x$ and making substitutions for the gradient and Hessian, we have

$$h_i(e;x) \approx U_i\left(w_{i0}^x\right) + U_i'(w_{i0}^x)\Delta A\left[e^T\left(\Pi_i^T Q A^+\right)^T\right] + \frac{1}{2}U_i''(w_{i0}^x)\Delta A^2\left[e^T\left(\Pi_i^T Q A^+\right)^T\left(\Pi_i^T Q A^+\right)e\right]. \tag{E-18}$$

Now, taking expections with respect to e, conditional on x

$$\mathbb{E}_{e|x}\left[h(e;x)|x\right] \approx U_i\left(w_{i0}^x\right) + \frac{1}{2}U_i''(w_{i0}^x)\Delta A^2 \cdot \mathbb{E}_{e|x}\left\{\left[e^T\left(\Pi_i^T Q A^+\right)^T\left(\Pi_i^T Q A^+\right)e\right]|x\right\}. \tag{E-19}$$

The first-order term drops out because $\mathbb{E}_{e|x}\left[e|x\right] = 0$ by assumption. Focusing on

the quadratic form in the final summand, standard results imply

$$\mathbb{E}_{e|x}\left\{\left[e^T\left(\Pi_i^T QA^+\right)^T\left(\Pi_i^T QA^+\right)e\right]|x\right\} = tr\left[\mathbb{E}_{e|x}\left[ee^T|x\right]\left(\Pi_i^T QA^+\right)^T\left(\Pi_i^T QA^+\right)\right] \tag{E-20}$$

$$= tr\left[\sigma_e^2\mathbb{I}\left(\Pi_i^T QA^+\right)^T\left(\Pi_i^T QA^+\right)\right] \tag{E-21}$$

$$= \sigma_e^2 tr\left[\left(\Pi_i^T QA^+\right)^T\left(\Pi_i^T QA^+\right)\right] \tag{E-22}$$

$$= \sigma_e^2||\Pi_i^T QA||_F^2. \tag{E-23}$$

The last expression is a basic property of the matrix Frobenius norm (Li et al. 2015).

Putting it all together, we have the following approximation to the expected utility for person i:

$$\mathbb{E}[U_i(W_i)] = \mathbb{E}_x\left[\mathbb{E}_{e|x}\left[h(e;x)|x\right]\right] \tag{E-24}$$

$$\approx \mathbb{E}_x\left[U_i\left(w_{i0}^x\right) + \frac{1}{2}U_i''(w_{i0}^x)\Delta A^2\sigma_e^2||\Pi_i^T QA||_F^2\right] \tag{E-25}$$

$$= \mathbb{E}_x\left[U_i\left(w_{i0}^x\right)\right] + \frac{1}{2}\mathbb{E}_x\left[U_i''(w_{i0}^x)\right]\Delta A^2\sigma_e^2||\Pi_i^T QA||_F^2. \tag{E-26}$$

Note that we have used the fact that A, Q, and Π_i^T are all independent of x.

From Theorem 2 the accuracy of the matrix mechanism is

$$I = -\sigma_e^2\left(\Delta A\right)^2\left\|QA^+\right\|_F^2. \tag{E-27}$$

We can therefore substitute accuracy, I, into the expression for expected utility

$$\mathbb{E}[U_i(W_i)] \approx \mathbb{E}_x\left[U_i\left(w_{i0}^x\right)\right] - \left\{\frac{1}{2}\mathbb{E}_x\left[U_i''(w_{i0}^x)\right]\frac{||\Pi_i^T QA||_F^2}{||QA||_F^2}\right\} \times I. \tag{E-28}$$

The expression above rationalizes a model for individual-specific data utility that is linear in accuracy, I: $v_i^{Data}(I) = a_i + b_i I$.

F Details of Legislative Redistricting Example

This appendix describes the legal background for the legislative redistricting example in Section VI.A. These properties of the SDL applied in the 2010 PL94-171 can be deduced from U.S. Census Bureau (2012, p. 7-6), as quoted in Appendix A, and the details provided in U.S. Census Bureau (2002), which also reveals that no privacy protection was given to the race and ethnicity tables in the 1990 redistricting data. The origin of the decision not to protect population and voting-age population counts is difficult to trace in the law. Public Law 105119, title II, 209, Nov. 26, 1997, 111 Stat. 2480, amended 13 U.S.C. Section 141 to provide that: "(h) ... In both the 2000 decennial census, and any dress rehearsal or other simulation made in preparation for the 2000 decennial census, the number of persons enumerated without using statistical methods must be publicly available for all levels of census geography which are being released by the Bureau of the Census for: (1) all data releases before January 1, 2001; (2) the data contained in the 2000 decennial census Public Law 94171 [amending this section] data file released for use in redistricting; (3) the Summary Tabulation File One (STF1) for the 2000 decennial census; and (4) the official populations of the States transmitted from the Secretary of Commerce through the President to the Clerk of the House used to reapportion the districts of the House among the States as a result of the 2000 decennial census. ... . (k) This section shall apply in fiscal year 1998 and succeeding fiscal years." http://www.law.cornell.edu/uscode/text/13 13 U.S. Code (1954). These amendments to Title 13 concerned the use of sampling to adjust the population counts within states, as is permitted even under current law. They gave standing to obtain a copy of population count data that were not adjusted by sampling, should the Census Bureau publish such data, which it did not do in 2000 nor 2010. Even so, only the reapportionment of the House of Representatives must be done without sampling adjustments (U.S. Supreme Court 1999). Sampling aside, other statistical methods, like edits and imputations, including whole-person substitutions, are routinely applied to the confidential enumeration data before any tabulations are made, including those used to reapportion the

House of Representatives. These methods were upheld in Utah v. Evans (U.S. Supreme Court 2002).

References

13 U.S. Code. 1954. "USC: Title 13 - Census Act."

Erlingsson, Úlfar, Vasyl Pihur, and Aleksandra Korolova. 2014. "RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response." *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security - CCS '14*, 1054–1067.

Kifer, Daniel, and Ashwin Machanavajjhala. 2012. "A rigorous and customizable framework for privacy." *Proceedings of the 31st symposium on Principles of Database Systems - PODS '12*, 77.

Li, Chao, Gerome Miklau, Michael Hay, Andrew McGregor, and Vibhor Rastogi. 2015. "The matrix mechanism: optimizing linear counting queries under differential privacy." *The VLDB Journal*, 24(6): 757–781.

U.S. Census Bureau. 2002. "Census Confidentiality and Privacy 1790 to 2002." https://www.census.gov/prod/2003pubs/conmono2.pdf, (Cited on March 22, 2018).

U.S. Census Bureau. 2012. "2010 Census Summary File1–Technical Documentation." Department of Commerce, Economics and Statistics Administration.

U.S. Supreme Court. 1999. "DEPARTMENT OF COMMERCE v. UNITED STATES HOUSE (98-404) No. 98—404, 11 F. Supp. 2d 76, appeal dismissed; No. 98—564, 19 F. Supp. 2d 543, affirmed." https://www.law.cornell.edu/supct/html/98-404.ZO.html, (Cited on March 26, 2018).

U.S. Supreme Court. 2002. "UTAH V. EVANS (01-714) 536 U.S. 452 182 F. Supp. 2d 1165, affirmed." *https://www.law.cornell.edu/supct/html/01-714.ZS.html*, (Cited on March 22, 2018).