ESSAYS ON DATA PRIVACY CHALLENGES THAT FEDERAL STATISICAL AGENCIES CONFRONT IN A DATA-RICH WORLD

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

William Nelson Sexton

May 2020

ESSAYS ON DATA PRIVACY CHALLENGES THAT FEDERAL STATISICAL AGENCIES CONFRONT IN A DATA-RICH WORLD

William Nelson Sexton, Ph.D.

Cornell University 2020

With vast databases at their disposal, private tech companies can compete with public statistical agencies to provide population statistics. However, private companies face different incentives to provide high-quality statistics and to protect the privacy of the people whose data are used. When both privacy protection and statistical accuracy are public goods, private providers tend to produce at least one suboptimally, but it is not clear which. In the first paper, we model a firm that publishes statistics under a guarantee of differential privacy. We prove that provision by the private firm results in inefficiently low data quality in this framework.

When Google or the U.S. Census Bureau publish detailed statistics on browsing habits or neighborhood characteristics, some privacy is lost for everybody while supplying public information. In the second paper, we assert that to date, economists have not focused on the privacy loss inherent in data publication. In their stead, these issues have been advanced almost exclusively by computer scientists who are primarily interested in technical problems associated with protecting privacy. Economists should join the discussion, first, to determine where to balance privacy protection against data quality; a social choice problem. Furthermore, economists must ensure new privacy models preserve the validity of public data for economic research.

Differential privacy is a mathematical tool for protecting the confidentiality of records belonging to individuals. One of the key premises of differential privacy is that *any* measurement based on the confidential data must be altered with carefully chosen random noise before publication. In the third paper, we consider a scenario where the deployment of differentially private disclosure limitation technologies by official statistical agencies may not always occur under ideal conditions. For instance, internal decisions or external requirements (e.g., legal or contractual obligations) may stipulate that certain statistics must be published *exactly*. Additionally, overlapping datasets may have already been published. In this paper, we explain (1) the semantics of algorithms that satisfy differential privacy, (2) how the semantics are affected by release of exact statistics (computed directly from the confidential data), (3) how to attribute responsibility for any resulting information leakage, (4) how to provide privacy semantics for the combined information leakage.

BIOGRAPHICAL SKETCH

William Sexton is a Research Mathematical Statistician at the U.S. Census Bureau, where he serves as a leader in the design and implementation of differentially private algorithms for the 2020 Decennial Census of Population and Housing. William is also a PhD candidate in Economics at Cornell University. He received his M.S. in Economics from Cornell University in 2016 where he worked as a research assistant with the Labor Dynamics Institute under the direction of John Abowd, Ian Schmutte, and Lars Vilhuber. William received his M.S. in Mathematics from Brigham Young University in 2014, where he studied combinatorial matrix theory under Wayne Barrett. He received his B.S. in Mathematics with a minor in Computer Science from Brigham Young University in 2012.

ACKNOWLEDGEMENTS

I owe a great debt to my co-authors. Without their contributions, these papers would not have come to fruition. The first and second papers are based on joint work with John Abowd, Ian Schmutte and Lars Vilhuber. Some of the second paper appeared in Abowd et al. (2019). John, Ian, and Lars introduced me to differential privacy and have been excellent mentors throughout my graduate school experience. I'm also grateful for their insights into how sound scientific research should be carried out. Thank you for dedicating so much of your time to my educational and professional growth! Thank you for welcoming me on as a research assistant at Cornell. That opportunity rests among the most pivotal moments in not just my graduate studies but also my life. John, you are an outstanding chair. Thank you for investing so much personal attention into my research and my work at Cornell and at Census. Lars, your support has been invaluable. Thank you for teaching me the virtues of version control and proper code and data curation as well as handling so many critical logistics issues for me. Ian, you've been such a vital member of my committee. Thank you for your patient guidance as I initially learned the ropes as a researcher. Every conversation is a delight. You've been both an invaluable mentor and a good friend.

The third paper is based on joint work with Robert Ashmead, Daniel Kifer, Philip Leclerc, Ashwin Machanavajjhala. Dan and Ashwin taught me many valuable lessons in algorithm design and the semantics of privacy. And, I spent countless hours debating the semantics of invariants with Phil and Robert.

I'd also like to thank committee members David Easley and Vitaly Shmatikov.

They have been phenomenal. Each has always been quick to respond and open with honest and constructive feedback whenever needed.

My Census colleagues have been very supportive. Aref Dajani deserves a special thanks for his encouragement and insistence on maintaining a decent work-life balance.

I thank my parents for always being there for me. Thank you for always being willing to listen and also for sharing your abundant wisdom with me. I thank my daughters, Emma and Bonnie, for being the best girls in the world and brightening every day. And, I thank my wife Annette for her love, moral support, and consistent motivation as well as for all she's sacrificed along the way to help me through to the end.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

SUBOPTIMAL PROVISION OF PRIVACY AND STATISTICAL ACCURACY WHEN THEY ARE PUBLIC GOODS

1.1 Introduction

Traditionally, statistical agencies have been charged with publishing summaries of data collected from the nation's citizens and businesses. Their data collection activities are expensive, and at risk of losing funding, despite an increasing demand for reliable data.[1] In this environment, one option is to augment, or replace, public statistical programs with information collected by private companies. Indeed, companies that aggregate personal data—e.g., Apple, Facebook, Google, Indeed, LinkedIn, Microsoft, Uber, Upwork—are under pressure to use their vast databases in the public interest. They are, often in collaboration with academic researchers, developing innovative data products like Google Trends (Choi and Varian 2012), the Billion Prices Project (Cavallo and Rigobon 2016), and the University of Michigan Social Media Job Loss Index (Antenucci et al. 2014). Clearly, the private sector is capable of producing innovative data products and could provide them competitively to the public.

Why are population statistics provided by public statistical agencies rather than private firms? There are a number of potential explanations, but in this paper we focus on inefficiencies in how private providers trade off data privacy and

[1] For evidence of increasing demand for Census data, see the discussion in Ruggles et al. (2019).

accuracy. Following the fundamental law of information recovery (Dinur and Nissim 2003), increasing the accuracy of published statistical summaries necessarily results in a loss of privacy for the data owners. This means statistical agencies must perform a balancing act. Published statistics should be as accurate as possible without revealing too much information about any single individual or business. When the benefits of more accurate population statistics and privacy losses are shared by all citizens, we show that private provision will result in inefficiently low levels of data accuracy and inefficiently high levels of privacy protection.

To establish suboptimality of the private provision of population statistics, we model the problem faced by a private data custodian who wants to sell population statistics. Our model extends Ghosh and Roth (2015, GR hereafter), who consider the problem of a data custodian, or producer, with legal possession of confidential data that was originally provided by data owners. The custodian wants to sell population statistics based on the confidential data to data users, who need the statistical summaries sold by the custodian to improve decision-making. We formalize the tradeoff between privacy and accuracy by assuming the custodian publishes using a differentially private mechanism. Operating this mechanism to publish statistics with a given level of accuracy requires the data owners to incur a known and quantifiable loss of privacy. Ghosh and Roth (2015) establish a minimum-cost method for purchasing privacy-loss rights from the data owners. Unlike Ghosh and Roth, who treat the demand for accuracy as exogenous, we assume an endogenous demand for data accuracy, and focus on its implications for the efficiency of private provision. The producer therefore balances the de-

mand for statistical accuracy against a demand for privacy protection. We model consumers who have heterogeneous preferences for the accuracy of the published statistical summaries, as well as for privacy protection. This formulation nests the more intuitive case in which the users of data are distinct from the population on whom data are collected.

Our model of data publication is based on *differential privacy* (Dwork 2006; Dwork et al. 2006; 2017), which has been adopted by tech companies like Google and Apple, as well as by the U.S. Census Bureau.[2] Differential privacy is an approach to publishing statistical summaries from confidential data sources that allows the publisher to make explicit, mathematically rigorous, statements about how much privacy—measured as a quantity—is lost with each publication. Furthermore, differentially private publications can induce an explicit positive relationship between accuracy of the published data and the amount of privacy loss.[3]

Crucially, we model both privacy protection and accuracy as public goods. Thus, it is not *a priori* obvious whether the private provider will provide too much or too little privacy protection. Data accuracy is a public good, since any consumer may access and use the published data without reducing its accuracy for some other consumer (it is non-rival) and no consumer can block another consumer's use (it is non-excludable). In plain English, all persons can learn and

[2]See, for example, Erlingsson et al. (2014), Differential Privacy Team (2017), and Abowd and Schmutte (2019) regarding applications of differential privacy at Google, Apple, and Census respectively.

[3]For a non-technical introduction to differential privacy, see Wood et al. (2018). See Heffetz and Ligett (2014) for an introduction targeted toward economists. For a more comprehensive treatment, see Dwork and Roth (2014).

benefit from the use of high-quality data by others, and they can also access those data directly themselves. They value what they learn. And they understand that what they learn is more useful if it is more accurate. Privacy protection is also a public good because all individuals in the database benefit from the same level of privacy protection embodied in the producer's data publication process, an implication of the Ghosh-Roth mechanism (non-rivalry in consumption for privacy protection).

We find that private provision results in suboptimally low data accuracy. As in Samuelson's classic model (Samuelson 1954), the external benefit of data accuracy to all consumers is not captured by the willingness-to-pay of the consumer with the greatest private value. By contrast, the demand for privacy protection is derived from the data provider's cost-minimization problem. The provider buys just enough data-use rights (privacy loss) to sell the data accuracy to the consumer with the highest valuation. All other consumers use the published data for free.[4]

While the suboptimality of private provision of public goods is well-understood (Spence 1975), modeling the origin and nature of suboptimality in the market for population statistics is not. Given the adoption of differential privacy by the U.S. Census Bureau, and the increasing demand for public data products from tech companies, it is important to consider how markets might, and might not, appropriately balance society's interests in privacy protection and data quality (Abowd and Schmutte 2019). Our paper is also broadly related to recent work

[4]Study of this case may be of special interest for some business-data collection for industries with a small number of dominant organizations.

in the economics of privacy (Acquisti et al. 2016; Heffetz and Ligett 2014; Goldfarb et al. 2015), which focuses on the role of privacy in facilitating the efficient use of customer data. Few papers have considered the economic tradeoff between privacy protection and data quality in the production of population statistics. Ghosh and Roth (2015), on which we build, and related papers in electronic commerce (Li et al. 2014) assume the demand for accuracy is exogenous. In some settings, this is appropriate—for example, when a company is mining its customer data for internal use. Publishing data summaries is, as we now show, another matter altogether.

1.2 Preliminary Concepts

This section provides our formal definitions of privacy and data accuracy. Our definitions are based on a computer science literature studying formal privacy, and so may be unfamiliar to economists. Our summary draws on several sources to which we refer the reader who is interested in more details (Hardt and Rothblum 2010; Dwork and Roth 2014; Wasserman and Zhou 2010; Heffetz and Ligett 2014; Abowd and Schmutte 2019). Our notation follows Dwork and Roth (2014).

We introduce the notion of differential privacy, which is key to understanding our analysis. Differentially private data publications do not allow an outsider to learn "too much" about any individual data record based on statistical summaries of the full database. For our purposes, this framework is useful because

differential privacy tells us, for any level of data accuracy, how much privacy loss an efficient provider must be willing to tolerate.

1.2.1 Databases and Queries

A data custodian (e.g., Facebook, Google, the U.S. Census Bureau) possesses a database, D. Think of D as a table in which each row represents information for a single individual and each column represents a single characteristic to be measured. The database D contains N rows. We assume all variables are discrete and finite-valued, but this is not restrictive since continuous data are always given discrete, finite representations when recorded on search logs, email contents, network features, social media posts, censuses, surveys, or administrative record systems.

The notion of a *neighboring database* is crucial for the definition of differential privacy. Differential privacy captures the idea that published output should not change "too much" based on a single data item. We say database D' is a neighboring database of D if D' can be obtained by modifying a single row in D. We denote by $\mathcal{D}$ the set of all admissible databases.[5]

Data users are interested in learning answers to a *database query*. A query is a function, $Q : \mathcal{D} \to \mathbb{R}^K$ that maps input databases, $D \in \mathcal{D}$ to a vector in $\mathbb{R}^K$.

[5]Formally, D and D' are neighbors if the ℓ_1-norm of the difference in their histogram representations is 2. See Appendix A.0.1.

The concept of a database query can admit standard population statistics, like subgroup counts, means, variances, and so on, but is much broader. The case we consider in this paper focuses on publication of a single-valued query, but note that in general the query answer $Q(D)$ may be vector-valued.

1.2.2 Query Release Mechanisms, Privacy, and Accuracy

The data curator operates a query release mechanism that provides answers to queries Q given a database D.

Definition 1 (Query Release Mechanism) Let $\mathcal{Q}$ be a set of admissible single-valued queries. A query release mechanism M is a random function $M : \mathcal{D} \times \mathcal{Q} \rightarrow \mathbb{R}$ whose inputs are a database D and a query Q. The mechanism output is a probabilistic response to the query. The probability of observing $B \subseteq \mathbb{R}$ is $\Pr[M(D,Q) \in B|D,Q]$, the conditional probability, given D and Q, that the published query answer is in $B \in \mathcal{B}$, where $\mathcal{B}$ are the measurable subsets of $\mathbb{R}$.

Differential Privacy

Our definitions of differential privacy and accuracy for the query release mechanism follow Dwork et al. (2006) and Kifer and Machanavajjhala (2011). [6]

[6]Following the setup in Ghosh and Roth (2015), we are using the variant of differential privacy now known as *bounded* differential privacy. This means that the total number of records in the confidential database, called N below, is publicly known.

Definition 2 (ε-differential privacy) Query release mechanism M satisfies ε-differential privacy if for $\varepsilon > 0$, for all pairs of neighboring databases D, D', all queries $Q \in \mathcal{Q}$, and all $B \in \mathcal{B}$

$$\Pr\left[M(D,Q) \in B | D, Q\right] \leqslant e^{\varepsilon} \Pr\left[M(D',Q) \in B | D', Q\right],$$

where $\mathcal{B}$ are the measurable subsets of $\mathbb{R}$, and the randomness in M is due exclusively to the mechanism and not the process generating the database D.

Accuracy

We next define our measure of accuracy. For any query, $Q \in \mathcal{Q}$, the query release mechanism returns an answer, a, that depends on the input database, the content of the query response, and the randomization induced by the query release mechanism.

Definition 3 ((α, β)-accuracy) Query release mechanism M satisfies (α, β)-accuracy if for $Q \in \mathcal{Q}$ and a output from $M(D, Q)$,

$$\Pr\Big(|a - Q(D)| \leqslant \alpha \;\Big|\; D, Q \Big) \;\geqslant\; 1 - \beta$$

where $a, Q(D) \in \mathbb{R}$.

This definition guarantees that the error in the answer provided by the mechanism is bounded above by α with probability $(1-\beta)$.[7] The probabilities in the definition of (α, β)-accuracy are induced by the query release mechanism.

[7] This definition also appears in a more general form in Gupta et al. (2012).

1.2.3 Example

To illustrate the problem stylized by the model, consider the following scenario.

Under Public Law 94-171, the U.S. Census Bureau publishes the number of individuals of Hispanic origin in each census block. Many blocks have small populations. Publishing the size of the Hispanic population without statistical disclosure limitation can lead to disclosure of the ethnicity of individuals in that block.

Framed in terms of the model, the database of interest, D, has one row for each person in a given block, and includes a binary indicator of their Hispanic origin. A neighboring database, D', has the same rows, but the Hispanic origin is changed for exactly one entry. The query of interest, $Q(D)$, is the proportion of individuals of Hispanic origin in the block. If the Census Bureau publishes the answer to $Q(D)$ exactly, then an attacker who knows the Hispanic origin of all but one individual can learn the origin of the remaining individual with certainty. If, instead, it publishes a noisy proportion under a differentially private query release mechanism, $M(D, Q)$, such an attacker will remain uncertain of the origin of the remaining individual. The question is whether the noisy proportion still measures the true proportion with sufficient accuracy.

In the preceding example, the attribute of interest is a binary indicator. This may seem to be a restrictive assumption. As we noted earlier, continuous data are generally given discrete, finite representations when recorded in databases. For instance, Public Law 94-171 also requires publication of the number of individu-

als in blocks and tracts, classified by age, a continuous feature. However, in the database D, age is discretized to single years, and for block-level tabulations, to year ranges. An 52-year old individual is thus recorded as the binary response to the query "age = 52". More generally, our model applies for publication of any predicate query that asks whether an individual's characteristics satisfy a set of binary conditions.

1.3 Private Provision of Population Statistics

In this section, we model a data provider selling public statistics produced according to a differentially private mechanism by purchasing rights to use records in an underlying confidential database. Since accuracy and privacy protection are both public goods, the consequences of private provision are theoretically ambiguous until further structure is placed on the model. Given the structure below, we prove that too little data accuracy and too much privacy will be supplied by a private provider compared to the social welfare maximizing solution.

1.3.1 Model Setup

Each of N private individuals possesses a single bit of information, b_i, that is already stored in a database maintained by a trusted curator.[8] In addition to their private information, each individual is endowed with income, y_i.

Individuals each consume one unit of the published statistic, which has accuracy I defined in terms of (α, β)-accuracy, that is $I = (1 - \alpha)$. Since I is a public good, all consumers enjoy the benefits of I, but each consumer is charged the market price p_I, to be determined within the model, for her "share" of I, which we denote I_i, and the balance of the public good, which we denote $I^{\tilde{}i}$ is paid for by the other consumers. Thus, $I = I_i + I^{\tilde{}i}$ for all consumers.

The preferences of consumer i are given by the indirect utility function

$$v_i\left(y_i, \varepsilon_i, I_i, I^{\tilde{}i}\right) = \ln y_i + p_\varepsilon \varepsilon_i - \gamma_i \varepsilon_i + \eta_i\left(I_i + I^{\tilde{}i}\right) - p_I I_i. \tag{1.1}$$

Equation (1.1) implies that preferences are quasilinear in data accuracy, I, privacy loss, ε_i, and log income, $\ln y_i$.[9] We incorporate income and accuracy in the utility function because they are required for the arguments in this section.

[8] Trusted curator can have a variety of meanings. We mean that the database is held by an entity, governmental or private, whose legal authority to hold the data is not challenged and whose physical data security is adequate to prevent privacy breaches due to theft of the confidential data themselves. We do not model how the trusted curator got possession of the data, but we do restrict all publications based on these data to use statistics produced by a query release mechanism that meets the same privacy and confidentiality constraints.

[9] In this section, we keep the description of preferences for data accuracy and privacy protection as close as possible to the GR specification. They allow for the possibility that algorithms exist that can provide differential privacy protection that varies with i; hence ε_i appears in equation (1.1). They subsequently prove that $\varepsilon_i = \varepsilon$ for all i in their Theorem 3.3.

The term p_ε is the common price per unit of privacy, also to be determined by the model. The receipt $p_\varepsilon \varepsilon_i$ represents the total payment an individual receives if her bit is used in an ε-differentially private mechanism. The individual's marginal preferences for data accuracy (a "good") and privacy loss (a "bad," really an input here), $(\eta_i, \gamma_i) > 0$, are not known to the data provider, but their population distributions are public information. Therefore, the mechanism for procuring privacy has to be individually rational and dominant-strategy truthful.

We do not include any explicit interaction between the publication of statistical data and the market for private goods. This assumption is not without consequence, and we make it to facilitate exposition of our key point: that data accuracy may be under-provided due to its public-good properties. Violations of privacy might affect the goods market through targeted advertising and price discrimination. The accuracy of public statistics may also spill over to the goods market by making firms more efficient. These are topics for future work.

1.3.2 The Cost of Producing Data Accuracy

A supplier of statistical information wants to produce an (α, β)-accurate estimate, $\hat{s}$, of the population statistic

$$s = \frac{1}{N} \sum_{i=1}^{N} b_i \tag{1.2}$$

i.e., a normalized query estimating the proportion of individuals with the property encoded in b_i. This property could be something highly sensitive, such as the

individual's citizenship status, sexual orientation, or whether she suffers from a particular health condition.

Theorems 3.1 and 3.3 in Ghosh and Roth (2015) prove that publishing the statistic

$$\hat{s} = \frac{1}{N}\left[\sum_{i=1}^{H} b_i + \frac{\alpha N}{2\left(1/2 + \ln\frac{1}{\beta}\right)} + \text{Lap}\left(\frac{1}{\varepsilon}\right)\right] \tag{1.3}$$

provides (α, β)-accuracy, and requires a privacy loss of $\varepsilon_i = \varepsilon = \frac{1/2+\ln(1/\beta)}{\alpha N}$ from $H = N - \frac{\alpha N}{1/2+\ln(1/\beta)}$ members of the population. $\text{Lap}\left(\frac{1}{\varepsilon}\right)$ represents a draw from the Laplace distribution with mean 0 and scale parameter $\frac{1}{\varepsilon}$.

Purchasing the data-use rights from the H least privacy-loving members of the population; *i.e.*, those with the smallest γ_i, is the minimum-cost, envy-free implementation mechanism (Ghosh and Roth 2015).[10] GR provide two mechanisms for implementing their VCG auction. We rely on their mechanism *MinCostAuction* and the properties given in their Proposition 4.5. See Appendix A.0.2 for additional details.[11]

We now derive the producer's problem of providing the statistic for a given level of data accuracy, I. If p_ε is the payment per unit of privacy loss, the total cost

[10] We note for completeness that the statistic $\hat{s}$, while computed on only H cases from the population of N, is evaluated relative to the population quantity s. GR use the same accuracy measure as we do; namely Definition 3 with a single query in the query set, although they assume $\beta = \frac{1}{3}$ throughout. We restrict the choice of β to $\beta < 1/(1+\sqrt{e})$; the threshold required by the proof technique in GR, Theorem 3.1. Statisticians often use mean squared error instead of the absolute error embodied in this definition. Nevertheless, the statistic $\hat{s}$ also trades-off bias and variance relative to the correct population statistic. The term $\alpha N/\left[2\left(1/2 + \ln\left(1/\beta\right)\right)\right]$ is a bias correction.

[11] Note that the result in equation (1.3) holds regardless of any correlation between privacy preferences and data values. That is, even if it is biased, the summary produced using data from those consumers with the lowest privacy preferences still satisfies the accuracy guarantee.

of production is $c(I) = p_\varepsilon H\varepsilon$, where the right-hand side terms can be defined in terms of I as follows. Using the arguments above, the producer must purchase from $H(I)$ consumers the right to use their data to compute $\hat{s}$. Then,

$$H(I) = N - \frac{(1-I)N}{1/2 + \ln(1/\beta)}. \tag{1.4}$$

Under the VCG mechanism, the price of privacy loss must be $p_\varepsilon = Q\left(\frac{H(I)}{N}\right)$, where Q is the quantile function with respect to the population distribution of privacy preferences, F_γ. The lowest price at which the fraction $\frac{H(I)}{N}$ of consumers do better by selling the right to use their bit, b_i, with $\varepsilon(I)$ units of differential privacy is p_ε. $H(I)$ is increasing in I. The total cost of producing I is

$$C^{VCG}(I) = Q\left(\frac{H(I)}{N}\right) H(I)\varepsilon(I), \tag{1.5}$$

where the production technology derived by GR implies

$$\varepsilon(I) = \frac{1/2 + \ln(1/\beta)}{(1-I)N}. \tag{1.6}$$

1.3.3 Example

The results of Ghosh and Roth (2015) hold even with an arbitrary correlation between privacy preferences and measured characteristics. Obtaining accuracy in the presence of potentially extreme selection bias makes it necessary to count almost everyone. In practice, the costs of publication are determined by the level of privacy-loss required, and the preferences of someone with extreme aversion to privacy loss.

Recall the example from Section 1.2.3, which involved computing the share of the Hispanic population in a census block. If an analyst requires $(0.2, 0.1)$-accuracy in the estimate, then she must purchase privacy loss of $\epsilon \approx 14/N$ from $H = 0.93N$ people, or 93 percent of the block population. Note that the required privacy loss, ϵ, vanishes as the block size N increases. To obtain a stronger guarantee of $(0.05, 0.05)$-accuracy would require purchasing $\epsilon \approx 70/N$ from $H = 0.99N$ people. A very weak guarantee of $(0.4, 1/3)$-accuracy only requires privacy loss of $\epsilon \approx 4/N$ from $H = 0.75N$ people.

1.4 Suboptimality of Private Provision

Suppose a private profit-maximizing, price-taking, firm sells $\hat{s}$ with accuracy (α, β), that is, with data accuracy I at price p_I. Then, profits $P(I)$ are

$$P(I) = p_I I - C^{VCG}(I).$$

If it sells at all, it will produce I to satisfy the first-order condition $P'\left(I^{VCG}\right) = 0$ implying

$$p_I = Q\left(\frac{H(I)}{N}\right) H(I)\varepsilon'(I) + \left[Q\left(\frac{H(I)}{N}\right) + Q'\left(\frac{H(I)}{N}\right)\left(\frac{H(I)}{N}\right)\right] H'(I)\varepsilon(I) \quad (1.7)$$

where the solution is evaluated at I^{VCG}.[12] The price of data accuracy is equal to the marginal cost of increasing the amount of privacy protection–data-use rights–

[12]The second order condition is $P''\left(I^{VCG}\right) < 0$, or $\frac{d^2 C^{VCG}(I)}{dI^2} > 0$. The only term in the second derivative of $C^{VCG}(I)$ that is not unambiguously positive is $\frac{H(I)H'(I)^2\varepsilon(I)}{N^2} Q''\left(\frac{H(I)}{N}\right)$. We assume that this term is dominated by the other, always positive, terms in the second deriva-

that must be purchased. There are two terms. The first term is the increment to marginal cost from increasing the amount each privacy-right seller must be paid because ε has been marginally increased, thus reducing privacy protection for all. The second term is the increment to marginal cost from increasing the number of people from whom data-use rights with privacy protection ε must be purchased. As long as the cost function is strictly increasing and convex, the existence and uniqueness of a solution is guaranteed.

1.4.1 Competitive Market Equilibrium

At market price p_I, consumer i's willingness to pay for data accuracy will be given by solving

$$\max_{I_i \geqslant 0} \eta_i \left(\tilde{I^{i}} + I_i \right) - p_I I_i \tag{1.8}$$

where $\tilde{I^{i}}$ is the amount of data accuracy provided from the payments by all other consumers, as noted above. Consumer i's willingness to pay is non-negative if, and only if, $\eta_i \geqslant p_I$; that is, if the marginal utility from increasing I exceeds the price. If there exists at least one consumer for whom $\eta_i \geqslant p_I$, then the solution to equation (1.7) is attained for $I^{VCG} > 0$.

We next show that there is only one such consumer. It is straightforward to

tive. Sufficient conditions are that $Q()$ is the quantile function from the log-normal distribution or the quantile function from a finite mixture of normals, and that $\frac{H(I)}{N}$ is sufficiently large; *e.g.*, large enough so that if $Q()$ is the quantile function from the $\ln N\left(\mu, \sigma^2\right)$ distribution, $Q^{*\prime\prime}\left(\frac{H(I)}{N}\right) + \sigma^2 Q^{*\prime}\left(\frac{H(I)}{N}\right)^2 \geqslant 0$, where $Q^*()$ is the standard normal quantile function.

verify that the consumers are playing a classic free-rider game (Mas-Colell et al. 1995, pp. 361-363). In the competitive equilibrium, the only person willing to pay for the public good is one with the maximum value of η_i. All others will purchase zero data accuracy but still consume the data accuracy purchased by this lone consumer. Specifically, the equilibrium price and data accuracy will satisfy

$$p_I = \bar{\eta} = \frac{dC^{VCG}\left(I^{VCG}\right)}{dI},$$

where $\bar{\eta}$ is the maximum value of η_i in the population–the taste for accuracy of the person who desires it the most. However, the Pareto optimal consumption of data accuracy, I^0, solves

$$\sum_{i=1}^{N} \eta_i = \frac{dC^{VCG}\left(I^0\right)}{dI}. \tag{1.9}$$

Marginal cost is positive, $\frac{dC^{VCG}\left(I^0\right)}{dI} > 0$, and $\sum_{i=1}^{N} \eta_i > \bar{\eta}$; therefore, data accuracy will be under-provided by a competitive supplier when data accuracy is a public good as long as marginal cost is increasing, which we prove below. More succinctly, $I^{VCG} < I^0$. Therefore, privacy protection must be over-provided, $\varepsilon^{VCG} < \varepsilon^0$, by equation (1.6).[13]

[13]The reader is reminded that a smaller ε implies more privacy protection. It is also worth commenting that in the GR formulation the single consumer with positive willingness to pay is the entity running the VCG auction. That person is buying data-use rights from the other consumers, computing the statistic for publication, then releasing the statistic so that all other consumers may use it. That is why we have modeled this as a public good. Our formulation is fully consistent with GR's scientist seeking data for a grant-supported publication.

1.4.2 The Price-discriminating Monopsonist Provider of Data accuracy

Now consider the problem of a single private data provider who produces $\hat{s}$ with accuracy (α, β) using the same technology as in equations (1.5) and (1.6). We now allow the producer to price-discriminate in the acquisition of data-use rights–that is, the private data-accuracy supplier is a price discriminating monopsonist. This relaxes the assumptions of the VCG mechanism in GR to allow for the unrealistic possibility that the data accuracy provider knows the population values of γ_i. They acknowledge this theoretical possibility when discussing the individual rationality and dominant-strategy truthful requirements of their mechanism. They reject it as unrealistic, and we agree. We are considering this possibility to show that even when the private data-accuracy provider is allowed to acquire data-use rights with a lower cost strategy than the VCG mechanism, data accuracy will still be under-provided.

The producer must decide how many data-use rights (and associated privacy loss ε, the same value for all i) to purchase from each member of the database, or, equivalently, how much to charge members of the database to opt out of participation in the mechanism for computing the statistic. (They cannot opt out of the database.) Let $\pi \in \{0,1\}^N$ be the participation vector. Using the Lindahl approach, let $p^L_{\varepsilon_i}$ be the price that satisfies, for each consumer i,

$$p^L_{\varepsilon_i} \leqslant \gamma_i, \text{with equality if } \pi_i = 1. \tag{1.10}$$

Equation (1.10) says that the Lindahl prices are those such that the choice of ε is exactly the value that each individual would optimally choose on her own. Even with our assumption of linear preferences, the Lindahl prices are unique for every consumer who participates in the mechanism for computing the statistic.

Given a target data accuracy of $I = (1 - \alpha)$, the producer's cost minimization problem is the linear program

$$C^L(I) = \min_{\pi} \left(\sum_{i=1}^{N} \pi_i p_{\varepsilon_i}^L \right) \varepsilon \tag{1.11}$$

subject to

$$\sum_{i=1}^{N} \pi_i = N - \frac{(1-I)N}{1/2 + \ln(1/\beta)} \text{ and } \varepsilon = \frac{1/2 + \ln(1/\beta)}{(1-I)N}.$$

The solution is for the producer to set $\pi_i = 1$ for the H members of the database with the smallest $p_{\varepsilon_i}^L$ and $\pi_i = 0$, otherwise. Note that if

$$\frac{dC^L(I)}{dI} < \frac{dC^{VCG}(I)}{dI}$$

for all I, which will be proven in Theorem 1, then the Lindahl purchaser of data-use rights will produce more data accuracy at any given price of data accuracy than the VCG purchaser.

By construction, the Lindahl solution satisfies the Pareto optimality criterion for data-use rights acquisition that

$$\sum_{i=1}^{N} \pi_i p_{\varepsilon_i}^L = \sum_{i=1}^{N} \pi_i \gamma_i. \tag{1.12}$$

Once again, the supplier implements the query response mechanism of equation (1.3) with $\frac{1/2+\ln(1/\beta)}{(1-I)N}$-differential privacy and $(1 - I, \beta)$-accuracy but pays each

consumer differently for her data-use right. Notice that equation (1.12) describes the Pareto optimal privacy loss whether or not one acknowledges that the privacy protection afforded by ε is non-rival, only partially excludable, and, therefore, also a public good.

To implement the Lindahl solution, the data producer must be able to exclude the bits, b_i, of specific individuals when computing the statistic, and must have perfect knowledge of the every marginal disutility γ_i of increasing ε. When this information is not available, the producer can, and will, implement the first-best allocation by choosing a price through the VCG auction mechanism used by GR.

For readers familiar with the data privacy literature, we note that the statement that technology is given by equations (1.5) and (1.6) means that the data custodian allows the producer to purchase data-use rights with accompanying privacy loss of $\varepsilon = \frac{1/2+\ln(1/\beta)}{(1-I)N}$ from $H(I)$ individuals for the sole purpose of computing $\hat{s}$ via the query response mechanism in equation (1.3) that is $\frac{1/2+\ln(1/\beta)}{(1-I)N}$-differentially private and achieves $(1-I,\beta)$-accuracy, which is exactly what Ghosh and Roth prove.

1.4.3 Proof of Suboptimality

Theorem 1 If preferences are given by equation (1.1), the query response mechanism satisfies equation (1.6) for ε-differential privacy with $(1-I,\beta)$-accuracy, cost functions satisfy (1.5) for the VCG mechanism, and (1.11) for the Lindahl mecha-

nism, the population distribution of γ is given by F_γ (bounded, absolutely continuous, everywhere differentiable, and with quantile function Q satisfying the conditions noted in Section 1.4), the population distribution of η has bounded support on $[0, \bar{\eta}]$, and the population in the database is represented as a continuum with measure function H (absolutely continuous, everywhere differentiable, and with total measure N) then $I^{VCG} < I^L$ and $I^{VCG} < I^0$, where I^0 is the Pareto optimal level of I solving equation (1.9), I^L is the privately-provided level when using the Lindahl mechanism to procure data-use rights and I^{VCG} is the privately-provided level when using the VCG procurement mechanism.

Proof. By construction, $F_\gamma(\gamma)$ is the distribution of Lindahl prices. Given a target error bound α, corresponding to data accuracy level $I = (1-\alpha)$, the private producer must procure data-use rights from the respondents in the confidential data with $\varepsilon(I)$ units of privacy protection from a measure of $H(I)$ individuals. Define

$$p_\varepsilon^\ell = Q\left(\frac{H(I)}{N}\right),$$

for $\ell = VCG, L$. Note that p_ε^ℓ is the disutility of privacy loss for the marginal participant in the VCG and Lindahl mechanisms, respectively. The total cost of producing $I = (1-\alpha)$ using the VCG mechanism is equation (1.5):

$$C^{VCG}(I) = Q\left(\frac{H(I)}{N}\right) H(I)\varepsilon(I).$$

while the total cost of implementing the Lindahl mechanism is equation (1.11):

$$C^L(I) = \left(N \int_0^{Q\left(\frac{H(I)}{N}\right)} \gamma dF_\gamma(\gamma)\right) \varepsilon(I).$$

Using integration by parts and the properties of the quantile function,

$$\begin{aligned} C^L(I) &= \left[Q\left(\frac{H(I)}{N}\right) F_\gamma\left(Q\left(\frac{H(I)}{N}\right)\right) - \int_0^{Q\left(\frac{H(I)}{N}\right)} F_\gamma(\gamma)d\gamma \right] N\varepsilon(I) \\ &= \left[Q\left(\frac{H(I)}{N}\right) H(I) - N\int_0^{Q\left(\frac{H(I)}{N}\right)} F_\gamma(\gamma)d\gamma \right] \varepsilon(I). \end{aligned}$$

Differentiating with respect to I,

$$\frac{dC^L(I)}{dI} = \left[Q\left(\frac{H(I)}{N}\right) H(I) - N\int_0^{Q\left(\frac{H(I)}{N}\right)} F_\gamma(\gamma)d\gamma \right] \varepsilon'(I) + Q\left(\frac{H(I)}{N}\right) H'(I)\varepsilon(I).$$

The corresponding expression for $C^{VCG}(I)$ is

$$\frac{dC^{VCG}(I)}{dI} = Q\left(\frac{H(I)}{N}\right) H(I)\varepsilon'(I) + \left[Q\left(\frac{H(I)}{N}\right) + Q'\left(\frac{H(I)}{N}\right)\frac{H(I)}{N} \right] H'(I)\varepsilon(I).$$

Comparison of the preceding marginal cost expressions establishes that $0 < \frac{dC^L(I)}{dI} < \frac{dC^{VCG}(I)}{dI}$ for all I, since $N\int_0^{Q\left(\frac{H(I)}{N}\right)} F_\gamma(\gamma)d\gamma > 0$, $\varepsilon'(I) > 0$, $H'(I) > 0$, and $Q'() > 0$. The results stated in the theorem follow by using the equilibrium price for the private market sale of I, which is p_I in equation (1.1),

$$p_I = \bar{\eta} = \frac{dC^L(I^L)}{dI} = \frac{dC^{VCG}(I^{VCG})}{dI}.$$

Hence, $I^{VCG} < I^L$, since $\frac{dC^L(I^L)}{dI} < \frac{dC^{VCG}(I^L)}{dI}$ and the conditions on Q imply that $\frac{d^2C^{VCG}(I)}{dI^2} > 0$. Likewise, $I^{VCG} < I^0$, since $\sum_{i=1}^N \eta_i > \bar{\eta}$, and $\frac{d^2C^{VCG}(I)}{dI^2} > 0$. ■

1.5 Conclusion

The concept of differential privacy allows a natural interpretation of privacy protection as a commodity over which individuals might have preferences. In many

important contexts, privacy protection and data accuracy are not purely private commodities. When both are public goods, the market allocations might not be optimal. The solution to the social planning problem that optimally provides both public goods–data accuracy and privacy protection–delivers more data accuracy, but less privacy protection, than the VCG mechanism for private-provision of data. The reason is that the VCG mechanism for procuring data-use rights ignores the public-good nature of the statistics that are published after a citizen sells the right to use her private data in those publications.

This matters because the demand for public data is greater than ever, while funding for statistical agencies has been relatively stagnant. It is increasingly likely that data users will turn to private companies to obtain the information they demand. Our results suggest that, while the policy debate has centered on regulating the privacy loss from this trend, it is also important to counterbalance the demand for privacy against the social value of reliable population statistics. Abowd and Schmutte (2019) propose a model for determining the optimal balance between privacy and accuracy in this social choice framework. More research, both theoretical and empirical, will help data users and policy makers navigate our modern data-rich environment.

The VCG mechanism also underprovides accuracy compared with the fictitious Lindahl mechnanism. We did not identify any relation between optimal provision and the Lindahl mechanism, suggesting our results are sensitive to the setup. Our model inherits some limitations (e.g., simple market structure) of the

original GR framework. We also inherit the positive aspects of their model. In particular, the results of GR's Theorems 3.1 and 3.3 are not tied to a specific auction mechanism.

In concluding, we point out possible extensions of our model. We have assumed a private data custodian must purchase privacy rights in order to use data in a published statistic. We make this assumption because it reflects the growing demand that companies be held accountable for the privacy of their customers' data through increasingly explicit means. However, if we assume companies may freely use their customers' data, then our conclusions will likely be reversed, with privacy being under-provided. To address this possibility, one might consider alternative specifications for the property rights over privacy loss. Similarly, our model assumes that private firms are limited to the single-buyer model in their ability to elicit payment from customers for data accuracy. One might explore alternative formulations of the demand side of the market, including the use of governmental organizations as the preference aggregators in making the purchase offer. This extension mirrors the original Ghosh-Roth motivation of a researcher spending grant money to buy the statistic, then placing the result in an open-access scientific journal. Finally, our model treats the data held by a trusted data curator as ground truth. Data users concerns for accuracy will likely extend beyond our model's treatment of differential privacy as the focal source of error in the underlying database. We need better accuracy measurements and tools for incorporating other sources of error from the data generation process (e.g., edit constraints, imputation, coverage error) into the social choice problem.

CHAPTER 2

WHY THE ECONOMICS PROFESSION MUST ACTIVELY PARTICIPATE IN THE PRIVACY PROTECTION DEBATE

2.1 Introduction

Privacy protection and scientific output are public goods. When Google displays search content clearly derivative of your recent online history or when the U.S. Census Bureau publishes geographically detailed demographic data clearly descriptive of your own neighborhood, some privacy is lost for everybody while supplying information that can be repeatedly re-used to increase utility.

Economists studying privacy have not focused on decisions about privacy loss inherent in the data publication process. These issues have recently been advanced almost exclusively by computer scientists who focus on technologies for increasing information quality while protecting privacy. Abowd and Schmutte (2019) showed that decisions about protecting privacy and making information public inherent in publishing data from confidential sources can be addressed using traditional social welfare analysis. This embeds the computer scientists' contributions into a framework that allows social scientists to contribute to the debate about safe methods for analyzing and publishing confidential data.

Economists rely heavily on designed data and administrative records from governmental agencies to do critical research. These studies are often done under

the supervision of a statistical agency exercising its dual mandate to disseminate information and to protect the privacy and confidentiality of respondent data. We have long recognized that there is tension between these mandates. Cryptographers established in the early 2000s that there is a hard limit to the amount of fully accurate information that can be published from any finite confidential database (Dinur and Nissim 2003)—a budget constraint stated in terms of confidential information leakage. New methods of confidentiality protection, known as formal privacy in computer science, quickly followed.

The implications of database reconstruction for the work of statistical agencies were largely unexplored before the U.S. Census Bureau announced its research program (Census Scientific Advisory Committee (CSAC) Meeting, September 2016) and its decision to implement differential privacy (Dwork et al. 2006), the leading variant of formal privacy models, for the 2020 Census of Population (CSAC Meeting, September 2017). The Commission on Evidence-based Policymaking (2017) also explicitly recommended that statistical agencies embrace privacy-enhancing data analysis methods.

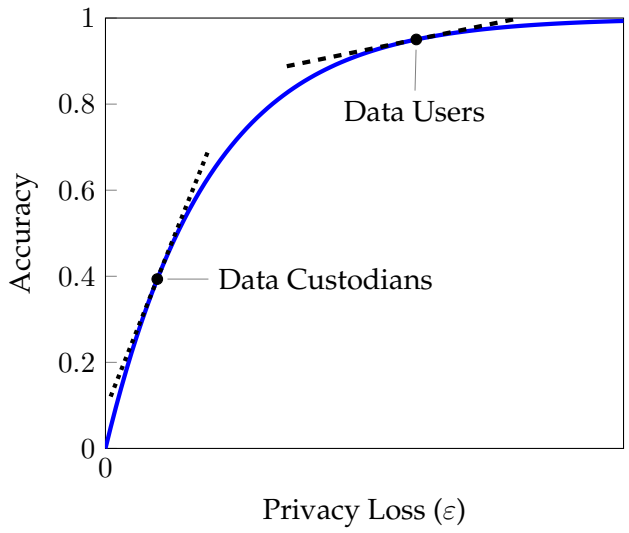

Figure 2.1: The trade-off between privacy loss and accuracy in data publication

These methods enforce an explicit trade-off between privacy protection and statistical accuracy, which economists will recognize as a production function. Implementation requires that the analyst acknowledge that fitting some models privately precludes fitting others unless more privacy loss is permitted. An explicit choice—outside the domain of computer science, but integral to economics—must be made: what is the optimal accuracy-privacy protection point for a given collection of data. The social choice is constrained by the formal privacy technology introduced by cryptographers. The preference mapping, on the other hand, must be expressed based on the uses of the published information and the attendant confidentiality risk. Figure 2.1 illustrates a typical production function with privacy loss (ε) on the x-axis and the accuracy of the data release on the y-axis. Accuracy is measured relative to releasing the data with no confidentiality protections

(accuracy = 1). Operating at a point beneath the production possibility frontier is inefficient whereas operating above the frontier is infeasible given the production technology. Technological innovation, a northwestern shift of the frontier, generally occurs with the discovery of new privacy mechanisms that more efficiently manage privacy loss. Such advancements are driven by research within the data privacy computer science community. Two different social welfare functions are illustrated. The tangent point labeled "Data Users" reflects the tendency of economists and other social scientists to favor accuracy over confidentiality protection. The point labeled "Data Custodians" reflects the tendency of data curators, often computer scientists, to favor privacy protection over accuracy. Privacy loss and information are inherently complements in production. Dissemination of increasingly accurate information cannot be accomplished without an accompanying increase in privacy loss as a by-product. On the other hand, a shift in social welfare that favors greater privacy must come at the expense of accuracy. Social scientists have behaved as if they could always have maximum accuracy in every published statistic. We must now re-design many of our analysis protocols to accommodate the constraints of provably effective privacy protection.

Economists are not the only ones. Apple (Differential Privacy Team 2017), Google (Erlingsson et al. 2014; McMahan and Andrew 2018), Microsoft (Ding et al. 2017), and many other information technology giants face the same conundrum. Because there are both technological and social preference components to the problem, ceding the debate to computer scientists focuses too much attention on the privacy mechanism and too little attention on how to do good social sci-

ence under a privacy-loss constraint. By drawing the attention of economists to their role in studying this problem, this paper begins to redress this imbalance.

2.2 Scientific Integrity Is the Highest Priority

Scientific discoveries are made by examining data using appropriate statistical techniques. We call those methods *inference-valid* when, under the maintained assumptions, the statistical conclusions have the probability distributions indicated by the theory. Inference-valid analyses allow the findings to generalize beyond the data from which they were derived. Scientists prefer to use the original, unmodified data as inputs, since any modifications may compromise the validity of the inference. However, when using the original data entails the risk of a breach of confidentiality, statistical disclosure limitation (SDL) is usually applied.

The value of SDL should not be measured merely as a function of its ability to protect against privacy loss, though this is surely important. Its value also lies in its ability to provide data that admit inference-valid analysis. Traditional SDL methods fail to uphold this principle (Abowd and Schmutte 2015). But inference-validity should be fully embodied in a modern SDL system, and formal privacy principles make this possible. These principles permit data curators to be completely transparent about the SDL system utilized. That is, every implementation detail, excluding the realizations of the "coin flips" of the mechanism, can be shared with data users without compromising the protection afforded by the

system. For example, the U.S. Census Bureau recently published the source code for its 2018 disclosure avoidance system (Census DAS Team 2019a;b). Google's release of tensorflow is another example (Google 2019). Transparency by data curators facilitates inference-valid analyses and therefore should be commended and encouraged.

2.3 The Roles to be Played by Economists

Amid the sea change in the way confidential data are made available for research, economists have three roles to play. As data users, we must gain a clearer understanding of what these changes mean for our ability to conduct valid research. The policy decisions made at statistical agencies have the potential to improve or further compromise inferential validity on any research question. Economists must be at the table as these decisions are made.

As data curators, we need to collaborate with the privacy community by engaging privacy experts in the design of our data publication systems as well as providing them with publication parameters often assumed to be exogenous in the differential privacy literature. Parameter settings that impact which models are best fit by a finite privacy-loss budget should not be left entirely up to the computer scientists. On the other hand, publication design choices that potentially constrain the design of efficient privacy technologies should not be made without a privacy expert present.

At a more fundamental level, economists can help guide policy-makers in deciding how to trade data accuracy off against privacy protection. The database reconstruction theorem implies that the information in a confidential database is finite. It can be allocated between the competing uses of protecting privacy or publishing more accurate statistics. This problem is in the economist's wheelhouse, particularly given that both uses are public goods.

Abowd and Schmutte (2019) describe this basic public choice problem, highlighting the key open areas for research. Fundamentally, we need to understand the social value of accessible, accurate data, and the social value of protecting the underlying confidential micro-data. Social scientists typically behave as if the social benefits of high-quality widely available data massively exceed the social costs of any associated privacy loss. This belief is not based on any rigorous theoretical or empirical evidence that we have found.[1] By contrast, cryptographers and other privacy experts tend to behave as if the social costs of privacy loss dwarf the benefits of data quality. To date, there are some models of the private demand for privacy (Ghosh and Roth 2015; Nissim et al. 2012), as well as a growing evidence base for the private costs of privacy loss (e.g. Acquisti et al. 2013).

[1]The literature on the value of public data is remarkably thin, notwithstanding early and important contribution of Spencer (1985), who developed a framework for modeling optimal data quality, and Panel on Statistics on Natural Gas (1985), who argued against the logical consistency of standard cost-benefit analysis for public data.

2.4 Traditional SDL Is Broken

Some resistance to the modernization of privacy protection arises from the mistaken belief that traditional SDL necessarily produces more reliable or even exact data with trivial re-identification risks (Ruggles et al. 2019). Newer methods are unfamiliar, while there are decades of research using data produced with traditional SDL. Researchers must replace general understanding of formal privacy with correctly reasoned comparisons of feasible alternatives.

It is important to realize that traditional SDL presents significant problems for social scientific research. Furthermore, the data demands imposed by quasi-experimental research designs exacerbate these flaws. The secrecy surrounding traditional SDL is a fatal flaw for social science. For example, when publishing micro-data, statistical agencies commonly swap records. The swap rate, the algorithm used to determine whether a record is at risk for swapping, and how the swapping is actually implemented, are all kept secret because there is no formal model to demonstrate that "enough" swapping was done. It might then be possible to undo the confidentiality protection afforded by the swapping (Abowd and Schmutte 2015).

Aside from the possible biases that swapping and other methods may introduce, traditional SDL introduces variability into the published data that should affect our inferences about what the underlying confidential data say about the world. This source of variability is almost never explicitly addressed in ensuring

that inferences based on SDL-protected data are valid. Even if we wanted to, because the details of traditional SDL are kept secret, it is usually not possible to account for it in estimation and inference.

Traditional SDL can also lead to bias in common research designs. Abowd and Schmutte (2015) show that current SDL practices introduce bias into estimates from linear regression models, instrumental variable models, and regression discontinuity studies. Analyses based on tabulated data, like the Quarterly Census of Employment and Wages (QCEW), are compromised by SDL rules that require cells influenced by just a few observations to be suppressed. The suppression rules are generally poorly documented for the same reasons as for swapping procedures, and in most studies, this suppression is nonignorable. Researchers have become comfortable with the practice of performing the analysis on the available data using the implicit assumption that suppressed data are missing at random. We should aspire to do better. We should aspire to procedures that are provably inference valid.

2.5 Formal Privacy Takes, but also Gives

A major concern regarding formal privacy systems is that they will change the ways in which researchers can access data, particularly micro-data. Exactly how formal privacy systems will affect the publication of detailed micro-data is the subject of extensive current research. Any change to the way published micro-

data are distorted is a matter of form and degree.

It is natural to mourn the loss of familiar data summaries, particularly as they may cause a break in continuity of data releases. But formal privacy methods also allow publishing new tabulations with far more detail than traditionally possible. Using input noise infusion, the Census Bureau publishes the Quarterly Workforce Indicators (QWI), county-industry level data on employment and job flows with demographic details and minimal suppression (Abowd et al. 2009). In the first official statistical publication using differential privacy, the Census Bureau publishes LEHD Origin-Destination Employment Statistics (LODES), complete block-level data on commuting flows (Machanavajjhala et al. 2008). The Post-Secondary Employment Outcomes (PSEO) pilot release (US Census Bureau 2018) relies on differential privacy to publish detailed earnings and employment outcomes for college and university graduates by degree level. Most recently, a team of Census Bureau and academics published the Opportunity Atlas (Chetty et al. 2018; Chetty and Friedman 2019), which provides inference-valid tract-level summaries of inter-generational mobility by race and gender—an outcome that is not feasible using traditional SDL.

2.6 Computer Scientists Are Right about Re-identification

The cryptographers found a fundamental defect in the approach statistical agencies have historically taken to SDL. The database reconstruction theorem shows

that it is always possible to reconstruct part or all of a confidential database using combinations of statistics published from that database. Therefore, even the publication of tabular summaries from, say, the decennial census or the American Community Survey is tantamount to a data security breach that releases all or part of the confidential database without intending to do so (e.g., block identifiers when only much more aggregated geography was planned). Every variable in the reconstructed micro-data is a potential identifier, even if the name and exact address cannot be reconstructed. Putting aside the legal and ethical questions of what constitutes a meaningful breach of privacy, it is fair to say that if we woke up tomorrow and learned that 50 percent of decennial census records, including detailed geography, had been exposed, we would find the statistical system under attack whether or not individuals could be re-identified from those released data. But they can, in fact be reidentified at substantial rates (Abowd 2019).

Differential privacy does not provide absolute protection against the disclosure of sensitive information. It trades absolute claims for relative ones, acknowledging at its core the impossibility of providing useful data summaries and complete privacy protection (Dwork et al. 2006). Formal methods control the global risk from reconstruction-abetted re-identification attacks using the privacy-loss budget ϵ. An adversary with auxiliary information that includes traditional identifiers (e.g., name and address) along with information that matches variables released via differential privacy, cannot improve the accuracy of any linkage for any person or any variable by more than a multiplicative factor of $e^{2\epsilon}$ (see Appendix B). If a statistical agency wants to provably limit linkage-based re-identification

attacks with a public degree of confidence, then it has no currently feasible choice except to adopt formal methods and stand by the privacy-loss budget it sets.

Traditional SDL also relies on uncertainty about whether a linkage-based attack produces a reliable re-identification. But agencies do not discuss the quantification of this risk—they do not release statistics on putative re-identifications (the number of records in the confidential database that their internal experiments were able to re-identify) nor on confirmed re-identifications (the number of putative re-identifications that were correct). If they did, one could discuss whether such a confirmation rate is acceptable. If a particular confirmation rate for re-identifications is acceptable, then formal methods can insure that the released data are consistent with a stated level of uncertainty about correct linkage re-identifications. For example, $\epsilon = 1.0$ guarantees that the improvement in the odds of a successful re-identification never exceeds $7.4 : 1$ for *any person* in the population when that person's data are used in the publications versus when they are deleted or replaced with an arbitrary record. An $\epsilon = 0.25$ guarantees that the improvement in the odds never exceeds $1.65 : 1$, and an $\epsilon = 0.1$ guarantees that the improvement never exceeds $1.2 : 1$. Many more examples of differential privacy's provable protection against re-identification can be found in Wood et al. (2018).

2.7 The Data Curator's Role in Algorithm Design

Algorithm design in differential privacy is the art of utilizing a privacy-loss budget as efficiently as possible to fit prescribed models as accurately as possible. Efficiency gains can occur along either dimension. That is, an inefficient production activity can be moved horizontally toward the frontier by discovering that the same accuracy may be achieved with lower privacy loss or the production point can be moved vertically if additional accuracy may be achieved at the same privacy loss. The former movement may highlight an improvement in privacy-loss accounting whereas the latter may result from an improvement in the allocation of the total budget among competing interests. Policy-makers need to be well-informed on the consequences of budget allocation choices to avoid unnecessarily diminishing data utility. Assuming the privacy analysis is done correctly and the budget is fully exhausted, algorithm design choices cannot improve the worst-case privacy-loss bound beyond the level implied by the theory. Thus data utility, rather than privacy considerations, is the driving factor behind algorithmic design decisions. Design changes are, in fact, production technology changes. Technological change affects the feasibility set of (ε, accuracy)-pairs as well as the location of the frontier. Additionally, a technological change often fundamentally alters the allocation problem, restricting, expanding, or otherwise morphing the set of computations that require a share of the total budget. Hence, the technology must be fixed before the allocation choice is made. That is, policy-makers cannot effectively carry out their role pertaining to the final budget and its allocation until

algorithm design is frozen.

In turn, algorithm design cannot be done effectively without a precise problem statement. The data curator must clearly set forth the solution criteria including the goal and requirements of the algorithm. We need to recognize the trade-offs associated with various solution criteria as well. We discuss three key aspects of problem specification.

The first is the specification of the data use-cases. The goal of data collection must be clearly expressed and the published data needs to be fit-for-use. To that end, the uses must be defined and importantly fitness-for-use must be measurable. Data utility metrics provide criteria for judging whether one privacy algorithm is better than another. This is also critical for judging whether one allocation of the privacy-loss budget outperforms another. However, sacrifices are unavoidable as there are in general no universally optimal privacy mechanisms (Brenner and Nissim 2014).

The second is the specification of the query workload. Although there are other publication paradigms, we focus attention on the scenario in which a statistical agency plans to publish a predefined set of tabulations. The query *workload* is a collection of queries of interest to the data curator. This collection is not necessarily identical to the set of tabulations intended for publication. For example, a data curator may plan to publish population counts for the set {total, non-voting age, voting-age, male voting-age, female voting-age} but this representation choice does not necessarily indicate a lack of interest in, say, the female non-voting age

population count.

A well-understood principle in privacy algorithm design is that directly answering the queries of interest is typically not the most efficient way to expend the privacy-loss budget (Li et al. 2010; McKenna et al. 2018). An indirect approach, taking measurements on an alternative set of *strategy* queries, yields better accuracy on the workload queries of interest provided the latter can be reconstructed from the former. Finding the optimal strategy is a complex problem. In practice, selecting a strategy is as much an art as it is a science. The effectiveness of using measurements on the selected strategy to reconstruct answers to the workload queries depends on the workload being correctly defined. A good strategy for a misspecified workload may be a bad strategy for the true queries of interest. Hence, the earlier distinction between queries of interest and the published data products. While strategy selection is a component of the algorithm design process, workload selection is ideally the responsibility of the data curator. Understanding the trade-offs between workloads is an area for future economic research.

The third is the specification of the privacy algorithm's input and output. Differential privacy is a property of an algorithm. It is not a specific algorithm nor is it a property of an algorithm's output. The output format can vary substantially from one algorithm to another (Dwork and Roth 2014). Existing publication systems may impose specific requirements on the output (e.g., micro-data, tabular summaries, maps, non-negative integer counts). It is a given that the output format must be suitable for the desired use-cases. However, rigid output require-

ments also restrict algorithm design, potentially diminishing the output's fitness-for-use.

For input specification, we focus on the *data universe*. The term data universe in differential privacy refers to the set of feasible database records. Specifications should identify infeasible attribute combinations, structural zeros. The data universe also relates to the concept of the unit of privacy. That is, the items that are protected by the algorithm whether they be businesses, or persons, or edges in a graph. Determining the unit of privacy is a fundamental part of algorithm development as it influences the impact a single item can have on the data analysis. Privacy experts call on subject matter experts to define specific components of some formal privacy frameworks (Kifer and Machanavajjhala 2012). Likewise, the unit of privacy decision does not fall outside the scope of a data curator's role. Other input considerations that can impact algorithm development range from data collection to data cleaning and imputation. Understanding trade-offs with different units of privacy or different data cleaning techniques, are also open research areas.

In each case, data curators can define these problem components in conjunction with the algorithm designers to promote statistical disclosure limitation methods that accommodate inference-valid research.

2.8 Moving Forward

To make progress, we should agree on the principles used to evaluate confidentiality protection mechanisms, whether traditional or formally private. Three components are essential.

First, agree on a *replication protocol* that confirms the provenance and authenticity of public-use inputs such as particular public-use data releases. Next, it identifies and confirms the provenance of the computations applied to those inputs to generate a specific set of outputs. Finally, the replication protocol confirms applying these computations to the public-use inputs produces the published outputs claimed in a particular scientific paper.

Second, agree on a *validation protocol* that confirms the provenance and authenticity of the confidential inputs used to produce the versions of the public-use inputs in the replication protocol. Next, it certifies the mapping from the computations applied in the replication protocol to the computations that must be applied to the confidential inputs to perform the same statistical analysis. Finally, the validation protocol produces outputs that are directly comparable to the outputs from the replication protocol.

Third, agree on a *comparison protocol*. Multiple candidate and historical public-use products may be put through the replication and validation protocols. The comparison protocol specifies how the validations will be compared, given that the replications are correct. Only the validations should be compared, because

these establish the properties of the scientific inferences, given the confidential data. There is no point in directly comparing replications from alternative inputs because such comparisons have no standard for correctness.

Ideally, an independent panel would conduct this process. However, such a panel would have difficulty vetting the validation protocol because curating the definitive versions of the confidential inputs to particular public-use products is very resource intensive. The Census Bureau's synthetic data program for the Survey of Income and Program Participation (SIPP) illustrates the commitment associated with maintaining replication and validation protocols (Benedetto et al. 2018).

Statistical agencies must commit resources to the research program outlined here. Professional organizations and curators of research data must be prepared to work with the agencies. Going forward, cooperation in achieving the objectives outlined in this section would position both the agencies and the research community to have increased confidence in the privacy protections and the scientific validity of all analyses based on the agencies' data.

CHAPTER 3

EFFECTIVE PRIVACY AFTER ADJUSTING FOR INVARIANTS, WITH APPLICATIONS TO THE 2020 CENSUS

3.1 Introduction

The Decennial Census of Population and Housing produces data that are used for apportioning the House of Representatives and redistricting all Congressional and state legislative bodies. The data are also used for intercensal population estimates, public policy, federal resource allocation, and general research. Historically, the Decennial Census has produced billions of statistics about a population whose size is measured in the hundreds of millions. Impossibility results, such as those of Dinur and Nissim (2003), imply that these statistics cannot be released too accurately without risking a reconstruction of almost all the confidential microdata from which the statistics are produced. To control the possibility and accuracy of such reconstruction attacks, the U.S. Census Bureau plans to deploy a differentially private disclosure avoidance system for the 2020 Census of Population and Housing. That system will allow published tables to be accompanied by formal privacy guarantees.

Differential privacy (Dwork et al. 2006), a notion of algorithm stability, provides a measure of information leakage and enables techniques for controlling the worst-case information leakage about any record. The bound on information

leakage is controlled by a parameter ϵ, known as the *privacy-loss budget*. The information leakage bound is determined by comparing the behavior of the algorithm between any pair of input databases that only differ in a single record. Attractive properties of differentially private algorithms include: (a) controlling the degradation of privacy loss over multiple releases,[1] and (b) limiting the ability of an attacker to *infer* sensitive properties of records even in the presence of side information (Kifer and Machanavajjhala 2012; 2014; Kasiviswanathan and Smith 2014; Wood et al. 2018). Differentially private algorithms require that any computation acting on the confidential database must be infused with random noise before being released. In particular, deterministic computations that use the confidential data in a non-trivial manner cannot be published in a differentially private system.

However, another design requirement of the disclosure avoidance system is that it accommodate the exact release of select statistics. Exact statistics computed from the confidential microdata are called *invariants*. For the 2018 End-to-End Census Test, the invariants originally considered necessary included the total population in each block, the voting-age population in each block, the number and type of group quarters facilities in each block, the number of housing units in each block, and the number of occupied housing units in each block. By definition, the number of occupied housing units is equal to the number of householders in the block.[2] The number of vacant housing units is equal to the number of housing

[1]Other statistical disclosure control techniques are susceptible to reconstruction of parts of the underlying microdata with as few as two releases (Ganta et al. 2008)

[2]The "householder" is "Person 1" on the Census questionnaire. This is usually the person who

units minus the number of occupied housing units. The non-voting-age population is equal to the total population minus the voting-age population. Thus, the totals for block-level householders, vacant housing units, and for the non-voting age population would also be invariants by construction. We initially viewed invariants as an externally mandated requirement to consider certain knowledge as public information; however, that viewpoint was not appropriately nuanced. This led the refinements that we now discuss.

We note that exact statistics about the confidential data can be obtained through alternative avenues. For example, external data sets and local knowledge can provide accurate information about the total population or number of housing units and group quarters facilities in many Census blocks (such information can include real-estate websites like Zillow and unstructured data like Google StreetView). Public facing Census operations such as the Local Update of Census Addresses (LUCA) operation and address canvassing can also leak accurate statistics about the confidential data. Non-differentially private publications released from overlapping internal Census data sets can reveal exact statistics as well.

Furthermore, interpretations of Constitutional and statutory requirements sometimes indicate that exact statistics, select unaltered tabulation of the confidential database, be released. For example, the U.S. Supreme Court (Department of Commerce v. United States House of Representatives, 1999) confirmed that the Census Act (13 U.S.C. Section 195) prohibits the use of "the statistical method

supplied the information for all persons living in the housing unit.

known as 'sampling' " for apportionment of the House of Representatives. Taken in concert with the one-person, one-vote rule (U.S. Supreme Court, Reynolds v. Sims, 1964), the redistricting amendments to Title 13 incorporated in PL94-171 (1975), and the amendments preceding the 2000 Census (Pub. L. 105–119, title II, § 209, Nov. 26, 1997, 111 Stat. 2480), "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."[3] In Utah v. Evans (2002) the U.S. Supreme Court clarified that other statistical methods, specifically including methods that change the number of persons in each state for the purposes of reapportionment, are not prohibited by 13 U.S.C. Section 195. Those methods are not "sampling."

However, the implications of these statutes and judicial rulings for statistical disclosure limitation are not clear and our understanding has evolved to view invariants as an internally imposed requirement. Historical practice at the Census Bureau has been to use disclosure limitation methods that do not change population totals at any geography (U.S. Census Bureau 2002). The theory underlying these methods originated in the 1970s (Fellegi 1972) when the formal privacy analysis of the implications of publishing exact statistics had not yet been discovered. When the form and extend of invariants is a matter of policy, the privacy implications should be carefully assessed by decision makers. The Data Stewardship and Executive Policy Committee (DSEP) is responsible for resolving policy matters related to privacy and confidentiality to ensure the Census Bureau upholds

[3] https://www.law.cornell.edu/uscode/pdf/lii_usc_TI_13.pdf, page 41

its commitment to protect the information of individuals.

The choice of invariants for the 2020 Census publications will be set by DSEP via policy, not engineering. These policies imply the Census Bureau must control all data publications (such as tabulations analogous to those appearing in PL94-171, Summary File 1 and Summary File 2(Bureau 2012)) in a manner that does not allow reconstructions of the protected data to exceed the bound implied by the global privacy-loss budget in the differentially private mechanism. The global privacy-loss accounting is materially changed by the presence of invariants. The presence of invariants amplifies the information leakage associated with any disclosure limitation system. To aid DSEP, we examine the privacy semantics of imposing various combinations of the previously defined invariants. We show that differential privacy allows the amplified information leakage from invariants to be controlled by appropriate management of the privacy-loss budget and acknowledgment that some risks from invariants cannot be quantified.

While differential privacy has many semantic interpretations, the main interpretation pursued in this document is now known as posterior-to-posterior semantics. It is based on ideas that originated in cryptography (Kasiviswanathan and Smith 2014; Wood et al. 2018). Intuitively, it compares an attacker's gain in inference arising from two scenarios – one in which a respondent truthfully reports a record and another in which the respondent reports a fictitious record. The posterior-to-posterior semantics of differential privacy holds in the presence of arbitrary side information. The generality of that claim admits the possibility of

counting invariants among the arbitrary side information an attacker possesses. However, the side information in question has a specific property of being an official publication explicitly released without disclosure limitation from confidential data under the stewardship of the Census Bureau. We have adopted the posterior-to-posterior sematics approach and properly specialized it to mechanisms that combine invariants with differential privacy.

In this paper, we first explain the semantics of differential privacy in the ideal setting without invariants. Next, we study the information leakage that results when invariants are added to the differentially private algorithms. We quantify upper bounds on the leakage due to the algorithms and the amplification due to the invariants. Finally, we apply these analyses to the invariants that are under consideration for the 2020 Census.

The introduction of invariants complicates the analysis of privacy. In extreme cases, invariants can be fully disclosive irrespective of whether any other publication occurs. By way of a simple example, consider a classroom setting in which a teacher reveals the exact number of students with passing grades. Any student with accurate prior knowledge of the class size can correctly deduce the number of failing students and also significantly improve their inference about whether any one individual is passing or failing. In a more stark case, suppose the teacher reveals exactly one student is passing. If you are that student, you would then know with certainty that every other student is failing. On the other hand, the idea that exactly releasing the total population of the United States, with no ac-

companying statistics, may lead to a privacy breach seems farfetched. However, the introduction of invariants is a catalyst that has the potential to greatly amplify the information leakage associated with disclosure algorithms. Suppose the 2020 Census disclosure avoidance system releases perturbed single-year age by sex distributions at the tract level. For this example, we ignore other age by sex tables historically published in finer geographic detail. Block level invariants that reveal the number of housing units and the number and type of group quarters facilities clearly amplify inferential disclosure risks. For example, knowledge that a particular block only contains an all-female juvenile correctional facility will certainly lead to an enhanced inference about any individual living there beyond what would have been inferred given only the tract level perturbed age by sex distribution. Thus, one can ask the following questions:

1. If there is a privacy breach, how much of it is due solely to the disclosure avoidance system? By analogy, how much is due to the released tract level sex by age tables? We analyze an upper bound on the privacy risk that can be attributed solely to the disclosure avoidance system. This style of analysis appears in Section 3.5.
2. How can one quantify both the leakage due to the disclosure avoidance system and the magnification effect due to invariants? By analogy, how can we quantify the extra risk due to the combination of the age by sex tables with the magnification from housing and group quarters invariants? One way is to compare the risk from two scenarios: (1) housing and group quarters information without sex by age tables, and (2) housing and group quarters information with sex by age tables. This style of analysis is discussed in Section 3.6 and Section 3.9.
3. What properties of the disclosure avoidance system can mitigate the amplified risk? The results presented in this paper show that differential privacy is one such property.

Our conclusions are the following:

- In the absence of invariants, differential privacy provides the strongest protection against database reconstruction among all accepted disclosure limitation techniques (which include k-anonymity, data swapping, cell suppression, ad hoc noise infusion, etc.). This conclusion continues to hold even when the privacy parameters and source code of the differentially private algorithm are released – such information does not degrade the protections provided by differential privacy.
- The privacy leakage due to the disclosure of the invariants is generally difficult to quantify using modern privacy analyses because it heavily depends on what an attacker knows (and there is no consensus about what is reasonable or unreasonable to assume). However, the main concern is how it amplifies the risks of subsequent data releases.
- The upper bound on privacy leakage solely attributed to algorithms satisfying differential privacy can be quantified and is the same as if there were no invariants.
- An upper bound for the amplified leakage can be obtained for the invariants under consideration for the 2020 Decennial Census. More precisely, different facts about a record are afforded different levels of protection, whereas without constraints, all facts about a record are protected equally well. This is a consequence of invariants – even simple invariants such as block-level population totals exhibit this behavior. Intuitively, they make it harder to protect the reported location of respondents but other information that cannot be inferred from location is better protected.

In Section 3.2 we describe the components of the Census of Population and Housing data, relying primarily on the schema that is used for the 2018 End-to-End Census Test. In Section 3.3, we present differential privacy and its properties in the absence of invariants. In Section 3.4, we briefly digress to summarize the privacy guidelines that led to this focus on differential privacy. In Section 3.5, we show how to quantify the information leakage, when invariants are present, that

is attributable to differentially private disclosure control algorithms. In Section 3.6, we develop results for analyzing the amplified privacy leakage of differentially private algorithms that is caused by the invariants. In Section 3.7, we apply these results to analyze the privacy loss amplification that could be caused by the invariants under consideration for the 2020 Census of Population and Housing. In Section 3.8, we generalize the posterior-to-posterior semantics of Section 3.6 to allow more nuanced privacy guarantees for situations which are not covered by the semantics of Section 3.6. In Section 3.9, we provide another interpretation of these posterior-to-posterior semantics in terms of odds ratios. With these refined semantics, we provide more nuanced guarantees for the proposed invariants in Section 3.10.

3.2 Characteristics of the Data

The Census collects information about a variety of different entities. These include individuals, housing units, households, group quarters facilities, and geographies.

3.2.1 Data Description

3.2.1.1 Geography.

The fundamental unit of geography is a block, which is used to create the full geographic hierarchy, including voting districts and other larger areas. Although geography is technically a lattice, its main structure is hierarchical:

- The entire United States
- 52 states and state-like entities, which include 50 States, the District of Columbia, and Puerto Rico.
- Counties and equivalent subdivisions (some states do not use the term "county")
- Tracts
- Block groups
- Blocks

Geographic information about blocks, such as whether they are fully or partially comprised of water, part of urban areas, part of rural areas, etc, are collected but not considered private. The block boundaries are designed in collaboration with the states and are based on knowledge of the residents in those regions. This complication is ignored and the analysis treats the boundaries as independent of the realized Census data. Enumeration geography is pre-specified, and so is fixed prior to the collection of realized data. Tabulation geography – which is the geography used in publications – is not pre-specified and may depend on results of the census-taking process. Tabulation geographies, and not enumeration geographies, are used in the disclosure avoidance system.

3.2.1.2 Group Quarters (GQ)

Group quarters are structures whose primary purpose is to house unrelated people. These include correctional institutions, military barracks, college dormitories, etc. Individuals residing in a group quarters type may be subject to Census edit rules, such as age restrictions imposed to correct perceived response errors (for instance, a minimum age restriction for residents of a nursing home would fix an error in which a resident of a Nursing home fills in an age of 3), and restrictions on sex (e.g., all-male, all-female, or unrestricted college dormitories). Such characteristics of group quarters include:

- Age restrictions
- Single-sex status (all-female, all-male, no sex restriction on residence)
- High-level type (institutionalized, non-institutionalized)
- 3 digit type code (see Table 3.1)
- Occupancy status (a GQ cannot be "vacant")

GQ Type 904 is considered extremely sensitive and is treated separately using methods that are confidential. There is an obligation to protect some characteristics of the group quarters. By definition, every group quarters facility must have at least one resident.

	INSTITUTIONAL GROUP QUARTERS
101	Federal Detention Centers
102	Federal Prisons
103	State Prisons
104	Local Jails and Other Municipal Confinement Facilities
105	Correctional Residential Facilities
106	Military Disciplinary Barracks and Jails
201	Group Homes for Juveniles (Non-Correctional)
202	Residential Treatment Centers (Non-Correctional)
203	Correctional Facilities Intended for Juveniles
301	Nursing Facilities/Skilled-Nursing Facilities
401	Mental (Psychiatric) Hospitals and Psychiatric Units in Other Hospitals
402	Hospitals With Patients Who Have No Usual Home Elsewhere
403	In-Patient Hospice Facilities
404	Military Treatment Facilities With Assigned Patients
405	Residential Schools for People With Disabilities
	NONINSTITUTIONAL GROUP QUARTERS
501	College/University Student Housing
601	Military Quarters
602	Military Ships
701	Emergency and Transitional Shelters (With Sleeping Facilities) for People Experiencing Homelessness
702	Soup Kitchens
704	Regularly Scheduled Mobile Food Vans
706	Targeted Non-Sheltered Outdoor Locations
801	Group Homes Intended for Adults
802	Residential Treatment Centers for Adults
900	Maritime/Merchant Vessels
901	Workers' Group Living Quarters and Job Corps Centers
903	Living Quarters for Victims of Natural Disasters
904	Religious Group Quarters and Domestic Violence Shelters

Table 3.1: Three-digit Group Quarters types Bureau (2012)

3.2.1.3 Housing units.

A housing unit is a structure where a family could (but does not have to) reside. The distinction between a housing unit and a group quarters can be subjective. A housing unit has a vacancy status (vacant or not) and a tenure (e.g., owned, rented, etc.). Every non-vacant housing unit must have a householder.

3.2.1.4 Housing status.

There is no official definition for a place where an individual might live. We define housing status as whether a respondent lives in a household and is a householder, or lives in a household and is not a householder, or lives in a group quarters.

3.2.1.5 Individuals

Every person resides either in a GQ or a household. Information collected about individuals includes:

- Sex
- Age
- Relation to householder (e.g., householder, spouse of, child of, parent of, step-child of, etc.). For people living in a group quarters facility, their relation to householder attribute is the facility's 3-digit GQ code.
- Hispanic origin: a binary attribute (yes or no). In the raw data, more detailed information (such as country of origin) is also recorded.

- Races. There are 6 major race categories and individuals can belong to any non-empty subset of them. We treat race as 6 binary attributes (with a structural zero – an impossible combination disallowed by edit rules – imposed to disallow the empty subset reply). In the raw data, more detailed information, such as country of origin, or hand-written race/tribe name is also recorded.

3.2.2 Cardinality of the Record Universe

All the above attributes, age, GQ Type, and block id, etc, are considered discrete and finite-valued. The set of all combinations of person-level attributes, the record universe, is a finite Cartesian product of finite attributes. Hence, the record universe has finite cardinality. In some contexts, the infeasible combinations, structural zeros, are omitted from the record universe. Of course, a subset of a finite set is still finite.

3.2.3 Input Data

The input data consist of the person-level records of every individual, including the code for the block in which their housing unit is located. Block-level records include counts of vacant and non-vacant housing units in each block and the characteristics of group quarters in that block. In the implementation for the 2018 End-to-End Census Test, the relationship of individuals to the householder is not tabulated, neither is sex, and age is coarsened to a binary indicator of voting age

status. The main reason for these simplifications is a matter of scope, the End-to-End Test was designed to focus on producing only a subset of the planned publication tables – the PL94-171 redistricting data – and not the full set of summary tabulations. In the End-to-End Test, given its scope, the full set of structural zeros is easily expressed; non-voting age individuals cannot reside in nursing facilities and individuals belong to at least one of the six major race categories. All other group quarters age restrictions overlap the voting age boundary. Single-sex restrictions are not in play. Most other structural zeros involve detailed age and/or relationship to householder – e.g., there cannot be more than a 50 year age gap between a householder and his or her spouse.

3.2.4 Output Data

The output ω is a set of records containing only attributes derivative of the attributes in the input data. We assume that all of the invariants that are present in the input data D can also be computed from ω. Specifically, we assume there is an algorithm Q that computes the values of the invariants. The input to Q can either be a database having the schema of D or of ω and the output ω is constrained so that $\mathrm{Q}(D) = \mathrm{Q}(\omega)$.

3.3 Privacy in the Ideal Setting

Differential privacy (Dwork et al. 2006) is a cryptographically-inspired privacy definition that is used to design *mechanisms* (i.e., algorithms that protect records in a database). It is designed to hide an individual's contribution to any published statistics. It is also designed to be transparent – the privacy parameters and the source code of differentially private mechanisms can be released without compromising privacy. In contrast, methods like data swapping become insecure when the algorithm source code and privacy parameters (e.g., the swapping rate) become public. We introduce the data protection semantics of differential privacy through the example of randomized response.

3.3.1 Randomized response and posterior-to-posterior guarantees

In the context of surveys, individuals have roughly three choices: to participate and provide truthful information, to participate but provide falsified information, or to not participate. Incentives for non-participation or falsification include the time commitment and privacy concerns about answering questions. The field of survey design can address the first disincentive. Disclosure avoidance technology can address the second. One of the earliest disclosure avoidance solutions was *randomized response* (Warner 1965), which predates differential privacy but is

also the first known algorithm that satisfies differential privacy. Under one common variations of this methodology, the randomized response protocol is such that a respondent flips a weighted coin (e.g., heads with probability 0.51) and is instructed to answer a yes/no question truthfully if the coin lands heads and otherwise to provide the opposite (false) answer (e.g., with probability 0.49). The odds that the respondent answered truthfully are $0.51/0.49 \approx 1.04$ meaning that the data collector has a large degree of uncertainty about any information of the respondent (odds of 1 equal perfect uncertainty).

Now imagine that a respondent is worried about her privacy and determines that she does not want to submit her data. Instead, she is given the option to submit a default answer (either "yes" or "no", but determined ahead of time) in place of her own. In this case she still flips the weighted coin to determine if she provides the default answer or its opposite. Because the respondent does not submit her data, this scenario can be considered *private* for her. Note that in order for the scenario to be considered *private* it is important that the default option be independent of her true data.

The importance of the *private* scenario is that we want to use it as a comparison to the scenario in which the respondent submits her true data to the randomized response protocol and show there is very little difference in the inference that can be made about her given the output. Consider one respondent – Respondent A – and suppose that Respondent A's truthful answer is "yes" and the default answer is "no". Consider the following two scenarios – one where Respondent A inputs

the truthful answer into the randomized response protocol and another where Respondent A inputs the default value into the randomized response protocol. Let E_t (respectively, E_d) denote the event that Respondent A decides to use the truthful answer (respectively, default value) into the randomized response protocol. Let O_y (respectively, O_n) be the event that the protocol produces the output "yes" (respectively, "no"). Then we see that:

$$\frac{P(O_y \mid E_t)}{P(O_y \mid E_d)} = \frac{0.51}{0.49} \approx 1.04$$
$$\frac{P(O_n \mid E_t)}{P(O_n \mid E_d)} = \frac{0.49}{0.51} \approx 0.96$$

In other words, the output distribution is almost the same, regardless of whether Respondent A uses the true record in the randomized response protocol or the default record. Therefore there is little incentive for Respondent A not to participate in the survey for privacy reasons because the output when she does participate is almost the same as the output in the *private* scenario in which she submits a default record. Note that if the default value was instead "yes", then the odds would both be equal to 1.

A Bayesian interpretation can be added on top of these semantics. Consider an attacker with a prior belief about Respondent A and the general population. This prior could incorporate side information – for instance, the attacker may be a neighbor and so might know the sex and age of A, and might also know statistical

information such as how age is correlated with how people answer the survey questions. We impose no restrictions on the attacker's prior. Let R be a variable representing Respondent A's attribute. Then simple calculations show:

$$\frac{0.49}{0.51} \leqslant \frac{P(R = \text{“}yes\text{''} \mid O_y, E_t)}{P(R = \text{“}yes\text{“} \mid O_y, E_d)} \leqslant \frac{0.51}{0.49}$$
$$\frac{0.49}{0.51} \leqslant \frac{P(R = \text{“}no\text{''} \mid O_n, E_t)}{P(R = \text{“}no\text{''} \mid O_n, E_d)} \leqslant \frac{0.51}{0.49}$$

In other words, no matter what the output is (O_y or O_n), the posterior inference about Respondent A is almost the same under the scenario in which the Respondent A participates and the *private* scenario in which she submits a default value as input to the randomized response protocol. This type of interpretation of privacy, comparing the distance between the posterior distribution in a scenario in which the respondent's data is used to the posterior distribution in a *private* scenario where it is not, is known as posterior-to-posterior semantics.

3.3.2 Differential Privacy

The development of differential privacy has demonstrated two important consequences. First, the privacy guarantee of randomized response can be extended to a wider variety of mechanisms when the data collector is trusted (i.e., when the respondents give their true responses to the data collector and the data collector runs the mechanism on the data). Second, differential privacy enables the use of mechanisms that are much more statistically efficient. For example, randomized response can be used to estimate the true number of "yes" answers among n re-

spondents with standard deviation proportional to $\sqrt{n}$. On the other hand, with differential privacy, this number can be estimated with standard deviation that is constant with respect to n.

There are many variants of differential privacy. We will work with the *bounded* ϵ-differential privacy definition, which assumes all databases have a fixed size n, because of its natural fit with the premise of a Census, to provide an exact enumeration of the U.S. population. Let X^n be the set of all possible input databases of size n.

Definition 4 (Bounded DP) A mechanism M satisfies *bounded ϵ-differential privacy* if for every pair of databases $D, D' \in X^n$ such that D and D' differ by the modification of one record, and every set of outputs $S \in range(M)$, we have:

$$Pr(M(D) \in S) \leqslant e^{\epsilon} Pr(M(D') \in S)$$

where the probability is only taken with respect to the randomness in M and not with respect to the data.

The parameter ϵ is known as the *privacy-loss budget*. Values of ϵ near zero represent more privacy whereas larger values represent greater risk of privacy loss. In the previous randomized response example, the privacy-loss budget was $\ln \frac{0.51}{0.49} \approx 0.04$. The qualifier *bounded* is used to emphasize that the requirement is over all pairs of datasets that have the same fixed number of records n, in contrast with *unbounded* differential privacy (Kifer and Machanavajjhala 2011).

In the posterior-to-posterior interpretation of differential privacy, we compare

the posterior inference of an arbitrary attacker about a respondent's record under two scenarios: (1) the *actual release* in which the respondent's true value is used in the mechanism M and (2) the *private counterfactual* in which the respondent's record is not used in the mechanism and instead replaced with a default value. If the mechanism satisfies the differential privacy definition, then the attacker's posterior inference can be shown to be bounded relative to the *private counterfactual*.

More specifically, let R_A be the variable representing respondent A's information, with r_A being the true value reported to the survey provider. Let D be distributed according to the joint distribution of the records in the database and $\pi(D = d)$ be an adversary's prior beliefs about that database defined over possible databases $d \in X^n$. That is, D is a random variable defined on the Borel measurable subsets of the sample space X^n. A realization of D is $d \in X^n$. The prior π may incorporate any side information about individuals in the data along with general statistical information about the population. Let D_{-A} represent the random variable derived from D when the response for Respondent A is removed from the data (or similarly d_{-A} for a fixed realization of the database d). If Respondent A's data R_A is used, then the database is $D = D_{-A} \cup \{R_A\}$. Instead, if Respondent A's record is replaced with the fixed default record, then the database is $D' = D_{-A} \cup \{r_d\}$ (or $d' = d_{-A} \cup \{r_d\}$ for a fixed realization d).

The hypothetical scenario in which Respondent A's data is deleted and replaced by a fixed default record r_d, to maintain the total, is a privacy baseline. This counterfactual is considered *private*, assuming r_d does not depend on r_A, in

the sense that the mechanism cannot act on Respondent A's information because it is excluded from the mechanism's input. The differentially private mechanism $M()$ is run on the data and the result ω is released. For clarity, we note the randomness of $M(D)$ captures the randomness in the mechanism $M()$ as well the attacker's uncertainty about the database D whereas $M(d)$ considers only the randomness about $M()$ for a fixed database d. In our model, the attacker knows the value of the default record and knows whether Respondent A's true record was omitted and replaced by the default record. Assume the attacker is interested in making inference about Respondent A, then for any possible value r of R_A, we can consider the ratio of the attacker's posterior inference if the true or default record is included in the database:

$$
\begin{aligned}
&\frac{\pi(R_A = r \mid M(D) = \omega)}{\pi(R_A = r \mid M(D') = \omega)} \qquad (3.1)\\
&= \frac{\sum\limits_{d \in X^n} \pi(R_A = r, D = d \mid M(D) = \omega)}{\sum\limits_{d \in X^n} \pi(R_A = r, D = d \mid M(D') = \omega)}\\
&= \frac{\sum\limits_{d \in X^n} \pi(R_A = r, D = d, M(D) = \omega) \Big/ \pi(M(D) = \omega)}{\sum\limits_{d \in X^n} \pi(R_A = r, D = d, M(D') = \omega) \Big/ \pi(M(D') = \omega)}\\
&= \frac{\sum\limits_{d \in X^n} \pi(R_A = r, M(d) = \omega, D = d) \Big/ \sum\limits_{d \in X^n} \pi(M(d) = \omega, D = d)}{\sum\limits_{d \in X^n} \pi(R_A = r, M(d') = \omega, D = d) \Big/ \sum\limits_{d \in X^n} \pi(M(d') = \omega, D = d)}\\
&= \frac{\sum\limits_{\substack{d \in X^n \\ d\, s.t. R_A = r}} \pi(D = d) Pr(M(d) = \omega) \Big/ \sum\limits_{d \in X^n} \pi(D = d) Pr(M(d) = \omega)}{\sum\limits_{\substack{d \in X^n \\ d\, s.t. R_A = r}} \pi(D = d) Pr(M(d') = \omega) \Big/ \sum\limits_{d \in X^n} \pi(D = d) Pr(M(d') = \omega)}\\
&= \frac{\sum\limits_{\substack{d \in X^n \\ d\, s.t. R_A = r}} \pi(D = d) Pr(M(d) = \omega) \Big/ \sum\limits_{\substack{d \in X^n \\ d\, s.t. R_A = r}} \pi(D = d) Pr(M(d') = \omega)}{\sum\limits_{d \in X^n} \pi(D = d) Pr(M(d) = \omega) \Big/ \sum\limits_{d \in X^n} \pi(D = d) Pr(M(d') = \omega)}\\
&\leqslant \frac{\sum\limits_{\substack{d \in X^n \\ d\, s.t. R_A = r}} \pi(D = d) e^{\epsilon} Pr(M(d') = \omega) \Big/ \sum\limits_{\substack{d \in X^n \\ d\, s.t. R_A = r}} \pi(D = d) Pr(M(d') = \omega)}{\sum\limits_{d \in X^n} \pi(D = d) e^{-\epsilon} Pr(M(d') = \omega) \Big/ \sum\limits_{d \in X^n} \pi(D = d) Pr(M(d') = \omega)}\\
&= \frac{e^{\epsilon}}{e^{-\epsilon}} = e^{2\epsilon}
\end{aligned}
$$

because differential privacy guarantees $e^{-\epsilon} P(M(d')) \leqslant P(M(d) = \omega) \leqslant e^{\epsilon} P(M(d') = \omega)$. We can similarly establish the ratio is bounded below by $e^{-2\epsilon}$. Thus for any value r, the attacker's posterior probability about Respondent A's record when his or her record is used is no more than $e^{2\epsilon}$ relative to the posterior

probability about Respondent A's record in the private counterfactual where Respondent A's record is replaced with a default value. In short, we say that use of the Respondent A's record sharpened inference by at most $e^{2\epsilon}$.

Another consequence of the differential privacy definition is that the relative posterior probability comparing any two values of respondent A's record is bounded by $e^{2\epsilon}$. Let r_1 and r_2 be any two possible values of R_A supported by the prior $\pi(D)$, then

$$\frac{\pi(R_A = r_1 \mid M(D) = \omega)/\pi(R_A = r_2 \mid M(D) = \omega)}{\pi(R_A = r_1 \mid M(D') = \omega)/\pi(R_A = r_2 \mid M(D') = \omega)} \in [e^{-2\epsilon}, e^{2\epsilon}] \tag{3.2}$$

See Appendix C for the proof. Equation 3.2 further implies that the inference that can be made about Respondent A's record is only improved by at most $e^{2\epsilon}$ relative to the private counterfacutal in which Respondent A's record is replaced with a default record. This is the same guarantee as for randomized response, but differential privacy can help obtain better statistical efficiency. For example, the number of "yes" answers in a dataset can be released with standard deviation $\sqrt{2}/\epsilon$ no matter the size of the dataset, by taking the true count and adding Laplace noise with scale $1/\epsilon$ (Dwork et al. 2006).

It is important to note that for an attacker who is making inferences about Respondent A, the prior $\pi(D)$ can be arbitrary. Thus, no matter what belief an attacker may have had about Respondent A, differential privacy guarantees that the posterior belief is not very much improved when utilizing Respondent A's data in the mechanism.

The term "privacy loss" will be used frequently hereafter. Unless otherwise specified, *privacy loss* informally relates to the posterior-to-posterior semantics interpretation of privacy loss as an expression comparing the posterior inference of an attacker under two scenarios. Although the posterior-to-posterior expressions studied throughout share a common form, the formal meaning of "privacy loss" varies from one expression to another. For narrative convenience, we abuse the terminology throughout. An interpretation of each new expression will be given as the need arises. Additionally, privacy loss generally references a worst-case bound taken over arbitrary attacker priors, all neighboring datasets, and all possible outputs. However, it may sometimes be used relatively as well. Privacy-loss budget always refers to the ϵ of differential privacy which measures the maximum difference of the log odds of observing any output across neighboring data sets.

3.4 Privacy Principles

The field of statistical disclosure limitation (SDL) is over 50 years old. One may ask why the focus here is on differential privacy rather than on older methods, such as data swapping, controlled rounding, etc., (Willenborg and de Waal 1996). The reason is the emergence of important criteria for SDL methods (Dwork et al. 2006; McSherry 2009; Kifer and Lin 2010).

First, mechanisms (i.e., algorithms for SDL) must be accompanied by a measure ℓ of information leakage, so that $\ell(M)$ is a quantification of how much in-

formation is leaked by M. Information leakage measures must be *closed under post-processing*. This criterion can be explained as follows. Suppose M_1 is any mechanism on D and M_2 is any mechanism on $M_1(D)$, i.e., $M_2(M_1(D))$, then $\ell(M_2) \leqslant \ell(M_1)$. In information theory, this is known as the information processing inequality. In the case of ϵ-differential privacy, the privacy-loss budget ϵ serves as a measure of information leakage.[4] It is well-known (e.g., McSherry (2009)) that this measure of information leakage is closed under post-processing.

The next criterion for leakage measures is *composition*. It can be explained as follows. Let D be a dataset. Let M_1 and M_2 be two mechanisms and let M_3 be the mechanism that returns $M_1(D)$ and $M_2(D)$, one could release: $M_1(D)$ only, $M_2(D)$ only, or both (i.e., $M_3(D)$). Clearly, the last case should be considered the most disclosive as the outputs of M_1 and M_2 together leak at least as much information about the confidential data as either output would alone. The composition property requires that their measured information leakage be subadditive: $\ell(M_3) \leqslant \ell(M_1) + \ell(M_2)$. This allows information leakage to be treated like a monetary cost: the cost of releasing $M_1(D)$ and $M_2(D)$ together is at most the sum of their individual costs. This property allows statistical agencies to develop disclosure avoidance algorithms from smaller pieces – if the total allowable information leakage is, say, at most 3, then one set of tabulations can be produced with a mechanism M_1 whose privacy leakage is measured as 1 and another set of tabulations can be produced with a mechanism M_2 whose privacy leakage is measured as 2.

[4]Specifically, ϵ is equal to $\sup_{\omega, D_1, D_2} \ln \frac{P(M(D_1)=\omega)}{P(M(D_2)=\omega)}$, where the supremum ranges over all pairs D_1, D_2 that differ on the value of one record.

The combined release of both tabulations therefore satisfies the desired bound of 3.

Many disclosure control methods do not satisfy composition. For example, two independent releases of information using k-anonymity can exactly reveal many records in the original data (Ganta et al. 2008).

The privacy-loss budget ϵ from differential privacy is one of the few known leakage measures that have the composition property and are closed under post-processing.

Next, one may ask why the the focus here is on comparing posterior distributions rather than on formulating a leakage measure that compares an attacker's prior distribution to the posterior distribution after seeing the output $M(D)$ of a privacy mechanism M. The main reason is that prior-to-posterior semantics rely on a consensus of what prior or set of priors are reasonable. Such a consensus is unlikely to happen with Census data.

3.5 Privacy Loss and Invariants

We now return to the problem of quantifying the privacy loss of a differentially private mechanism in the presence of invariants. Consider the case of two algorithms Q and M. Algorithm Q operates deterministically on the collected data and outputs the values I of the invariants. Recall that invariants are a set of sta-

tistical tabulations that are released exactly without formal privacy protections. The algorithm M satisfies bounded ϵ-differential privacy and produces an output ω. We assume the algorithmic details (e.g., source code) of M and Q are public knowledge.

3.5.1 Generally Unquantifiable Aspect of the Privacy Loss

One of differential privacy's selling points is that it offers broad guarantees against arbitrary attacker priors. However, when any invariant is released, this claim no longer holds. There is unavoidable privacy loss that cannot be quantified without making assumptions about an attacker's prior. Therefore it is impossible to bound the privacy loss due to invariants for arbitrary attacker priors in the same multiplicitive manner as before. Invariants provide information that will allow certain attackers to completely rule out certain record types for a respondent, or in the more extreme case, have complete certainty about one or more attributes of a respondent. To see this, consider an arbitrary attacker's prior about the records in the database $\pi(D)$. Marginalizing $\pi(D)$ to Respondent A, leads us to the attacker's prior about Respondent A, $\pi(R_A)$. After releasing the invariants $Q(D) = I$, the updated prior for Respondent A would be $\pi(R_A|Q(D) = I)$, and defining invariants and priors such that for some record value r, $\pi(R_A = r) > 0$ but $\pi(R_A = r|Q(D) = I) = 0$, is a straightforward exercise. In the extreme case, releasing the invariants gives the attacker certainty about the value of Respondent A's record, meaning $\pi(R_A = r) < 1$ and $\pi(R_A = r|Q(D) = I) = 1$.

Instead considering posterior-to-posterior semantics, we run into a similar problem when comparing $\pi(R_A = r|Q(D) = I)$ with the attacker's inference in a counterfactual in which a default record was used, in place of Respondent A's record, to compute the invariants $\pi(R_A = r|Q(D') = I')$. The invariants $Q(D) = I$ and $Q(D') = I'$ can conflict such that certain record values for Respondent A are ruled out in one scenario but not the other. Thus the attacker would have significantly different inferences between the actual release and the private counterfactual (i.e., the posterior-to-posterior comparison can be unbounded). As a result, a general multiplicative bound is not possible and we view the initial privacy loss due to the release of invariants as generally unquantifiable.

One may question the realism of such examples. This is a fair concern as the conclusions of an attacker can vary greatly, depending on the priors used. Yet, the alternative practice of enumerating classes of "reasonable" attackers is knowingly flawed. The conundrum before us is that in considering the release invariants, policy makers must weigh an unquantifiable privacy risk against whatever benefits may be derived from releasing certain exact statistics.

3.5.2 Privacy Loss due to the Mechanism given the Invariants

One way to view privacy loss in the presence of invariants is to only consider the privacy loss conditional on the invariant information, effectively treating the true invariants as public information. More specifically, the private counterfac-

tual does not omit Respondent A's record from the invariant computation (i.e., $Q(D)$ is released), but then replaces his or her record with a default value r_d for the execution of M. Thus $M(D')$ is published, where $D' = D_{-A} \cup r_d$. In this counterfactual world, the attacker knows that, after the invariants were computed, the true record of Respondent A was replaced with r_d. Thus r_d must be determined independently of the true record r_A. In order to quantify the privacy loss due to the mechanism M given the invariants, the natural posterior-posterior comparison is:

1. *Actual Release*: $\mathrm{Q}(D)$ and $M(D)$.
2. *Private Counterfactual*: $\mathrm{Q}(D)$ and $M(D')$

Following the same steps as in Section 3.3, we bound the posterior-to-posterior ratio:

$$\frac{\pi(R_A = r \mid \mathrm{Q}(D) = I, M(D) = \omega)}{\pi(R_A = r \mid \mathrm{Q}(D) = I, M(D') = \omega)} \in \left[e^{-2\epsilon}, e^{2\epsilon}\right] \tag{3.3}$$

Therefore conditioning on the release of invariants I, the privacy loss due to the mechanism M is at most $e^{2\epsilon}$. While this result is accurate, it does not acknowledge the full impact of the invariants on the mechanism. To illustrate this in a simple way, consider the following artificial example. Suppose each respondent provides an integer from the range $[0, k]$. Thus the data D consist of n integers $R_1, \ldots, R_n$. Consider the following invariants: $R_1 + R_2$, $R_2 + R_3$, $R_3 + R_4, \ldots, R_{n-1} + R_n$. There are $n - 1$ linear equations on n unknowns, so that knowing R_j for any j means all of the other R_i can be determined exactly.[5]

[5]Note that in some situations, the invariants exactly determine all the data. For example, if even one sum $R_i + R_{i+1}$ equals 0, the invariants reveal every record.

An example of a mechanism M that satisfies bounded ϵ-differential privacy is one that adds independent Laplace random variables Z_i with scale k/ϵ (and variance $2k^2/\epsilon^2$) to each R_i, where the constant k is determined by the mathematical properties of the queries that M is permitted to accept. (Dwork et al. 2006).

Example 2 In the absence of invariants, the only available estimate of R_1 is the output $R_1 + Z_1$. The mean of this estimate is R_1 (i.e., it is unbiased) and its variance is $2k^2/\epsilon^2$ (the variance of the Laplace random variable Z_1).

Example 3 However, the invariants do exist and they do affect information leakage. Consider the actual case where M uses the true value of R_1. Because of the invariants there are now many ways to get independent estimates of R_1.

- The noisy estimate $R_1 + Z_1$ provides an estimate of R_1 with variance $2k^2/\epsilon^2$
- The noisy estimate $R_2 + Z_2$ provides another independent noisy estimate of R_1 (obtained by subtracting $R_2 + Z_2$ from the invariant $R_1 + R_2$) with variance $2k^2/\epsilon^2$.
- In a similar manner, each $R_j + Z_j$ provides yet another independent noisy estimate of R_1 with variance $2k^2/\epsilon^2$

These n noisy estimates can be averaged to obtain a new estimate of R_1 that has variance (i.e. squared error) equal to $2k^2/(n\epsilon^2)$. Note that use of the constraints $0 \leqslant R_i \leqslant k$ has the potential of reducing variance even more. For example, if we know that $R_1 + R_2 = 1$, then R_1 is either 0 or 1.

Example 4 Now suppose R_1 is changed to k before running M. Then M does not use the true value of R_1 in its computation. Now we have the following estimates:

- The noisy estimate $k + Z_1$ provides an (inaccurate) estimate of R_1 with squared error at most $k^2 + 2k^2/\epsilon^2$ (which would be the case if the true value was $R_1 = 0$). Call this estimate Y_1.
- The noisy estimate $R_2 + Z_2$ provides an independent unbiased noisy estimate of R_1 (obtained by subtracting $R_2 + Z_2$ from the invariant $R_1 + R_2$) with variance $2k^2/\epsilon^2$. Call this estimate Y_2.
- In a similar manner, each $R_j + Z_j$ provides yet another independent unbiased noisy estimate of R_1 with variance $2k^2/\epsilon^2$. We use Y_j to refer to the resulting estimates.

These n noisy estimates can be averaged to obtain a new estimate of R_1 that has squared error equal to

$$\begin{aligned} E\left[\left(R_1 - \frac{1}{n}\sum_{i=1}^{n} Y_i\right)^2\right] &= E\left[\left(\frac{1}{n}\sum_{i=1}^{n}(R_1 - Y_i)\right)^2\right] \\ &= \frac{E[(R_1 - Y_1)^2] + \sum_{i=2}^{n} E[(R_i - Y_i)^2]}{n^2} \\ &\leqslant \frac{k^2 + 2k^2/\epsilon^2 + (n-1)2k^2/\epsilon^2}{n^2} \\ &= \frac{k^2}{n^2} + 2k^2/(n\epsilon^2) \end{aligned}$$

Thus feeding M an incorrect value can increase the variance by at most k^2/n^2. When n is large, this increase in variance is negligible and the overall squared error of the estimate of R is small regardless of whether M used the real value of R_1 or not (and further refinements to this inference, such as using the constraints that $0 \leqslant R_i \leqslant k$ would serve to reduce the variance even more). Hence the invariants greatly amplify the information leakage.

The reason for this sharp inference is that the invariants in this example induced a strong dependence between all of the records. Thus, while the resulting privacy breach is mostly due to the invariants (rather than M), the possibility of reconstruction cannot be ignored and it is desirable to have semantics for the amplified privacy leakage caused by Q possibly interacting unfavorably with M.

In Definition 4, note the importance of neighboring databases is to ensure an attacker has difficulty distinguishing between similar databases given the output of the mechanism, and thus remains uncertain about the true database. However, when invariants are released the attacker can rule out any database that is not consistent with the invariants, decreasing the uncertainty. Further, the number of neighboring databases, as defined by the modification of one record, may be significantly reduced because some modifications may not be consistent with the invariants. Even releasing the total database size reduces the number of possible databases. In extreme cases, there may not be any consistent neighboring databases, rendering the privacy bound $Pr(M(d) \in S) \leqslant e^{\epsilon} Pr(M(d') \in S)$ meaningless. In order to better account for the impact of invariants on neighboring databases of the mechanism, we introduce the idea of *modification strategies* and *group modification strategies*.

3.6 Tools for Analyzing Leakage Guarantees

For a disclosure avoidance system that publishes invariants and differentially private output from the same input data, we consider three types of privacy loss:

1. The privacy loss due to publication of invariants. As discussed in Section 3.5, this loss may not be quantifiable without making assumptions about the prior of an attacker.
2. The privacy loss due to the mechanism M. As discussed in Section 3.5, this loss can be quantified using the comparison of inferences about a person's data in two hypothetical situations: one in which person A's data is used in the computation of M versus one in which person A's data was not used in the computation of M.
3. The privacy loss due to the composition of the invariant release Q and the mechanism M. This privacy loss is a result of dependence between information released by Q and M. It consists of loss attributed to M and the amplification of this loss due to Q (as illustrated in Examples 3 and 4 in Section 3.5).

In this section, we develop the tools and concepts that will help provide semantics for the privacy loss of Item (3).

To analyze, we need to consider the following general setting for data release:

- The invariant computation $\mathrm{Q}(D)$ produces its output I prior to the execution of the bounded ϵ differentially private algorithm M.
- The inputs of the algorithm M are a vector I of numbers and a dataset D. The output ω of $M(I, D)$ must satisfy $\mathrm{Q}(\omega) = I$. That is, we can always recover the first input of M from its output ω.
- The mechanism M satisfies bounded ϵ-differential privacy with respect to D – that is, for any fixed I and for all pairs (D, D') that differ on the value of

one record, and all ω, $P(M(I, D) = \omega) \leqslant e^{\epsilon} P(M(I, D') = \omega)$.[6]

After I and ω are published, the inferences that are possible depend on how they were generated. Thus a first attempt to quantify the privacy loss in Item (3) would be to compare posterior probabilities for the following two scenarios:

1. *Actual Release*: $\mathrm{Q}(D) = I, M(I, D) = \omega$ – all algorithms operate on the original data.
2. *Private Counterfactual*: $\mathrm{Q}(D_{-A} \cup \{r_d\}) = I', M(I', D_{-A} \cup \{r_d\}) = \omega$ – all algorithms operate on the modified data that uses a default record for Respondent A.

However, unless $I = I'$, one of the two events must occur with zero probability because $\mathrm{Q}(\omega)$ can only yield one of the two possible invariant inputs to M. As a result, if $I \neq I'$, no meaningful bound on the ratio of the posteriors can be provided (as in the discussion of Section 3.5 about the initial privacy loss caused by invariants). Thus we must develop more nuanced models of the private counterfactual that enforce $I = I'$.

To do so, we use the notion of a *modification strategy* (Bassily et al. 2013). Intuitively, one can think of a modification strategy as a procedure that scrubs information from a dataset about a person but preserves the invariants (i.e., delete

[6]Note that from the point of view of M, I is just a constraint on the output ω such that $Q(\omega) = I$. Privacy-loss amplification occurs when $Q(D) = I$, that is when I is the actual invariant of the data and not just a set of numbers specified *a priori*. This notation makes privacy-loss amplification easier to see, as the full algorithm (which uses the true invariants) is $M(Q(D), D)$.

record r_A and replace it with a record r_d such that $D_{-A} \cup r_d$ satisfies the invariants I). We define modification strategies as follows.

Definition 5 (Modification Strategy) A modification strategy ϕ_A is a (possibly randomized) modification of a database D that preserves the invariants $Q(D) = I$. The modified database is denoted as $\phi_A(D_{-A}, I)$ and ϕ_A satisfies the following conditions:

1. The data sets D and $\phi_A(D_{-A}, I)$ have the same number of records, and they only differ in a Respondent A's record, r_A, versus the replacement record, r_d. That is, $\phi_A(D_{-A}, I) = D_{-A} \cup r_d$
2. The equality $\mathrm{Q}(\phi_A(D_{-A}, I)) = \mathrm{Q}(D)$ holds for all D – the modification strategy maintains consistency with invariants.

Note that the inputs to the modification strategy ϕ_A are D_{-A} and I, so ϕ_A receives no information about the reported record of Respondent A except for information necessary to compute the invariant I. In this sense, the selection of r_d has minimal dependence on r_A.

For example, suppose the total population and voting-age population are released as invariants at the block level. In that case, comparing D_{-A} to I will completely determine the block and voting-age status of Respondent A (but not the specific age, or reported race, etc.). So ϕ_A will have access to information about the reported voting-age status and block, but will not have access to any other reported information from Respondent A. Under the modification strategy, it is

as if Respondent A only provided information about block and voting-age status and opted not to provide any additional information.

Under the posterior-to-posterior approach we compare the actual release to a counterfactual in which the modification strategy is used. With this idea, we consider the following two scenarios:

1. *Actual Release*: $Q(D) = I, M(I, D) = \omega$.
2. *Private Counterfactual*: ϕ_A is used everywhere, $Q(\phi_A(D_{-A}, I)) = I, M(I, \phi_A(D_{-A}, I)) = \omega$.

The privacy in the counterfactual is conditional on the invariants: no information from Respondent A is used beyond what is necessary to compute the invariants. This is a relative guarantee: the privacy loss due to invariants is generally not quantifiable, but any privacy loss beyond that is protected by the modification strategy. For example, if the invariants released are highly disclosive about Respondent A, then this respondent does not have much privacy to begin with. An extreme case is if the only record that could be added to D_{-A} in order to satisfy the invariants I is Respondent A's true record r_A (see Examples 3 and 4), then the invariants are highly disclosive and there is nothing left for the modification strategy to protect about respondent A. However, if the invariants are less disclosive – for example, if at least one variable in the data is not involved in the computation of the invariants – then there always exists a non-trivial modification strategy that can scrub information about these variable(s).

The posterior-to-posterior semantics are now used to analyze the degree to which a modification strategy will affect an attacker's inference:

$$\frac{\pi(R_A = r \mid \mathrm{Q}(D) = I, M(I, D) = \omega)}{\pi(R_A = r \mid \mathrm{Q}(\phi_A(D_{-A}, I)) = I, M(I, \phi_A(D_{-A}, I)) = \omega)} \tag{3.4}$$

where $\pi(D)$ is the attacker's prior beliefs about the records in the database.

Theorem 5 Let D be a random variable distributed according to the joint distribution of the records in the database including Respondent A, $\pi(D = d)$ be an arbitrary prior defined for $d \in X^n$ held by an arbitrary attacker, Q be a deterministic algorithm for computing invariants, M be a bounded ϵ-differentially private mechanism that takes both invariants and a database as input. Let R_A be the record for Respondent A, r be a possible record value, and ϕ_A be a modification strategy. Then for fixed I and any ω

$$\frac{\pi(R_A = r \mid \mathrm{Q}(D) = I, M(I, D) = \omega)}{\pi(R_A = r \mid \mathrm{Q}(\phi_A(D_{-A}, I)) = I, M(I, \phi_A(D_{-A}, I)) = \omega)} \in \left[e^{-2\epsilon}, e^{2\epsilon}\right].$$

See Appendix C for the proof. Theorem 5 implies that the inference an attacker can make about respondent A after the mechanism's release is at best $e^{2\epsilon}$ times sharper in actuality than in the private counterfactual. This brings together the two components of the privacy guarantee: (1) the counterfactual, which only utilizes information from Respondent A necessary to compute the invariants and can be considered private when the invariants are not themselves too disclosive, and (2) for small ϵ, attacker inference will be similar in the actual release and in the counterfactual.

Regardless of the particular modification strategy in use, the respondent should be "almost indifferent" between the actual release and the private counterfactual. This idea can be formalized by following the line of utilitarian reasoning employed by Ghosh and Roth (2015). Suppose $u_A : \mathcal{X} \rightarrow \mathbb{R}$ is a utility representation of Respondent A's preferences over some set of future events $\mathcal{X}$. Since two mechanisms release information, it may be natural to assume any distribution induced on $\mathcal{X}$ should condition on the outputs from each (i.e., consider an arbitrary $f : Range(\mathrm{Q}) \times Range(M) \rightarrow \Delta(\mathcal{X})$). However, by construction the output of Q can be fully recovered from the output of M, so it suffices to consider $g : Range(M) \rightarrow \Delta(\mathcal{X})$ (e.g., for an arbitrary f, we could define $g(M(I,D)) \equiv f(\mathrm{Q}(M(I,D)), M(I,D)) = f(I, M(I,D))$ where $\mathrm{Q}(D) = I$). If Respondent A faces a choice between participating in the actual release scenario and the counterfactual scenario implied by a particular φ_A, he or she can be assured that future expected utility from participation in the actual release scenario will decrease by at most a multiplicative factor of e^{ϵ} compared to the expected utility from opting-for the counterfactual.

$$\begin{aligned}
\mathbb{E}_{x \sim g(M(Q(D),D))}[u_A(x)] &= \sum_{x \in \mathcal{X}} u_A(x) \underset{g(M(Q(D),D))}{Pr}[x] \\
&\leqslant e^{\epsilon} \sum_{x \in \mathcal{X}} u_A(x) \underset{g(M(Q(\varphi_A(D)),\varphi_A(D)))}{Pr}[x] \\
&= e^{\epsilon} \mathbb{E}_{x \sim g(M(Q(\varphi_A(D)),D))} u_A(x)
\end{aligned}$$

$$\frac{\mathbb{E}_{x\sim g(M(\mathrm{Q}(D),D))}[u_A(x)]}{\mathbb{E}_{x\sim g(M(\mathrm{Q}(\varphi_A(D_{-A},I)),\varphi_A(D_{-A},I)))}[u_A(x)]} = \frac{\sum_{x\in\mathcal{X}} u_A(x) \underset{g(M(Q(D),D))}{Pr}[x]}{\sum_{x\in\mathcal{X}} u_A(x) \underset{g(M(Q(\varphi_A(D_{-A},I)),\varphi_A(D_{-A},I)))}{Pr}[x]} \in [e^{-\epsilon}, e^{\epsilon}]$$

The bound follows since $Pr_{g(M(Q(D),D))}[x]/Pr_{g(M(Q(\varphi_A(D_A,I)),\varphi_A(D_{-A},I)))}[x] \in [e^{-\epsilon}, e^{\epsilon}]$ as $g \circ M$ is bounded ϵ-differentially private. Note that since $M(I, D)$ is bounded ϵ-differentially private only with respect to its second input D, the invocation of the post-processing claim that $g \circ M$ is bounded ϵ-differentially private is only admissible because $Q(D) = Q(\varphi_A(D_{-A}, I))$.

3.7 Interpreting Guarantees for Census Invariants

In this section, we consider the privacy impact of the specific invariants under consideration by the U.S. Census Bureau for the 2018 End-to-End Census Test and the 2020 Census. For illustration, assume a data schema with the following variables:

1. Geography
2. Relationship To Householder (Including GQ status)
3. Sex
4. Age
5. Hispanic or Latino Ethnicity
6. Race

The list of invariants for consideration is as follows:

- **C1: Total population per block.**
- **C2: Voting-age population per block.**
- **C3: Number of housing units per block.**
- **C4: Number of occupied housing units per block.**[7]
- **C5: Number of group quarters facilities by type, per block.** The count of group quarters facilities by type per block includes:
  - Group quarters type (e.g., federal prison, college dormitory).
  - Single-sex institution status (e.g., female-only, male-only, unrestricted)
  - Age restrictions (e.g., minor-only, adult-only, unrestricted)

Given a choice of invariants from **C1** - **C5**, we consider the privacy implications of releasing the invariants along with the outputs of a mechanism satisfying differential privacy. We consider the posterior-to-posterior protections offered to each respondent according to Theorem 5 relative to the private counterfactual in which the respondent reports the minimal information necessary to compute the invariants.

Case 1: **C1** only

The only invariant is the total population per block. In any reconstruction, the block-level record count will always be correct. Per Theorem 5, the inference an attacker can make about a respondent after the mechanism's release is at best $e^{2\epsilon}$ times better than if all the attributes of a respondent's record, except the block,

[7]Note that this is equivalent to the number of householders per block.

were replaced with arbitrary values. The privacy protection for a respondent's block-level geocode is not quantified by Theorem 5.

Case 2: **C1, C2**

The invariants are the total population per block and the voting age population per block. In any reconstruction, the total population and the voting-age population at the block-level will always be correct. Per Theorem 5, the inference an attacker can make about a respondent after the mechanism's release is at best $e^{2\epsilon}$ times better than if all the attributes of a respondent's record, except the block and voting age status, were replaced with arbitrary values. This means than attributes like sex, race, and ethnicity can be modified arbitrarily. However, age can only be modified as long as the voting-age status does not change. The privacy protection for a respondent's block-level geocode or voting age status is not quantified by Theorem 5.

Case 3: **C1, C3**

The invariants are the total population per block and the number of housing units per block. In any reconstruction, the total population and the housing-unit counts at the block-level will always be correct. Recall that we define housing status as whether a respondent lives in a household and is a householder, or lives in a household and is not a householder, or lives in a group quarters. Per Theorem 5,

the inference an attacker can make about a respondent after the mechanism's release is at best $e^{2\epsilon}$ times better than if all the attributes of a respondent's record, except the block and housing status, were replaced with arbitrary values. Modifications to the relationship to householder attribute (from which housing status is derived) are constrained as follows:

1. For respondents in blocks with no housing units, no record for a person in a GQ can be altered to instead be in a household.
2. For respondents in blocks with at least one housing unit, there are no restrictions.

The privacy protection for a respondent's block-level geocode is not quantified by Theorem 5. The housing status for respondents in group quarters with no housing units is also not quantified by Theorem 5.

Case 4: **C1, C4**

The invariants are the total population per block and number of occupied housing units per block. In any reconstruction, the total population and the occupied housing-unit counts at the block-level will always be correct. Per Theorem 5, the inference an attacker can make about a respondent after the mechanism's release is at best $e^{2\epsilon}$ times better than if all the attributes of a respondent's record, except the block and housing status, were replaced with arbitrary values. Modifications to the relationship to householder attribute (from which housing status is

derived) are constrained as follows.

1. A non-householder respondent's record may be altered to be a GQ record (or vice versa)
2. A GQ or non-householder record cannot be altered to a householder record.
3. The relation to householder of a householder cannot be modified.
4. Respondents in blocks with only GQ units cannot have their record altered to instead be in a household.

The privacy protection for a respondent's block-level geocode is not quantified by Theorem 5. The housing status for householders in any block as well as for GQ residents in blocks with no housing units is not quantified by Theorem 5.

Case 5: **C1, C5**

The invariants are the total population per block and the number of group quarters facilities by type per block. In any reconstruction, the total population and the group quarters' facilities counts by type at the block-level will always be correct. Per Theorem 5, the inference an attacker can make about a respondent after the mechanism's release is at best $e^{2\epsilon}$ times better than if race and ethnicity were modified arbitrarily, while modifications to relation to householder, sex, and age satisfy the following restrictions.

1. Vacant group quarters are not tabulated. Individuals in group quarters containing only one person cannot change their relation to householder (i.e.,

they cannot change their housing status). If this group quarters has restrictions on age and sex then these attributes cannot be modified as well.

2. For respondents living in blocks containing no group quarters, the relation to householder can be modified to any value that is valid for people living in households.

The privacy protection for a respondent's block-level geocode is not quantifiable under Theorem 5. For individuals living in a group quarters containing only one person, the protections for relation to householder, age, and sex are also not quantified by Theorem 5.

Case 6: **C3, C5**

The invariants are the number of housing units and the number and type of group quarters facilities in each block. Per Theorem 5, the privacy protections vary for different individuals.

1. For individuals living in a group quarters that contains only one person, the inference an attacker can make about a respondent after the mechanism's release is at best $e^{2\epsilon}$ times better than if all attributes except block, relation-to-householder, age, and sex were arbitrarily modified. Sex can be arbitrarily modified if the group quarters does not contain restrictions on sex, and age can be modified in any way that is consistent with the age restrictions in the group quarters.

2. For any other individual, the inference an attacker can make about a respondent after the mechanism's release is at best $e^{2\epsilon}$ times better than if the entire record is modified arbitrarily subject to the following restrictions: if the reported block is changed to a block with no households, the relation to householder cannot imply that the person lives in a household. The altered age and sex must be consistent with the allowable age and sex for group quarters in that block. If the reported block is changed to a block with no group quarters, the relation to householder cannot imply the person lives in a group quarters.

The protections for block-level geocode, relation to householder, age, and sex of respondents in group quarters containing only one person are not quantified by Theorem 5.

Case 7: **C1, C3, C5**

The invariants are the population total at each block, the number of housing units per block, and the number and type of group quarters facilities per block. The privacy guarantee combines the restrictions taken from the privacy guarantees of Cases 4-6:

1. For respondents living in blocks that contain both GQ and housing units, a non-householder respondent's record may be altered to be a GQ record (and vice versa).
2. A GQ or non-householder record cannot be altered to be a householder record.

3. The relation to householder of a householder cannot be modified.
4. Respondents in blocks with only GQ units cannot have their record altered to instead be in a household.
5. For respondents living in blocks containing no group quarters, relation to householder can be modified to any value that is valid for people living in households.
6. For individuals living in a group quarters that contains only one person, all attributes except block, relation-to-householder, age, and sex can be arbitrarily modified. Sex can be arbitrarily modified if the group quarters does not contain restrictions on sex, and age can be modified in any way that is consistent with the age restrictions in the group quarters.

Case 8: **C1, C2, C3, C4, C5**

The invariants are the population total at each block, the voting-age population per block, the number of housing units per block, the number of occupied housing units per block, and the number and type of group quarters per block. The privacy guarantee combines the restrictions taken from each of the prior Cases 1-7. In particular, the inference an attacker can make about a respondent after the mechanism's release is at best $e^{2\epsilon}$ times better than if block and voting-age status were unmodified, race and ethnicity were modified arbitrarily, and the rest of the attributes modified as allowed by the restrictions of Cases 1-7 (in particular, individuals living in blocks with no households have the most restrictions).

3.8 Group Modification Strategies

Modification strategies are designed to reason about privacy for one person at a time and do so by filling in the deleted record of a respondent with values that preserved the invariants in the resulting data. Their ability to generate semantics is limited for some attributes (such as block, when block population totals are invariant).

It is possible to extend the idea of modification strategies and private counterfactuals to reason about groups of people and provide more nuanced privacy semantics. In this extension, the records of a set $\mathcal{S}$ of respondents are first removed, and then a modification strategy fills in those missing records in order to preserve the invariants. We call this a *group modification strategy* $\varphi_{\mathcal{S}}$.

Definition 6 (Group Modification Strategy) Given a set $\mathcal{S}$ of respondents, a group modification strategy $\varphi_{\mathcal{S}}$ is a (possibly randomized) modification of a database D that preserves invariants $Q(D) = I$. The modified database is denoted as $\varphi_{\mathcal{S}}(D_{-\mathcal{S}}, I)$ and satisfies the following conditions:

1. The data sets D and $\varphi_{\mathcal{S}}(D_{-\mathcal{S}}, I)$ have the same number of records and only differ in at most $|\mathcal{S}|$ records, those of the respondents in $\mathcal{S}$.
2. The equality $Q(\varphi_{\mathcal{S}}(D_{-\mathcal{S}}, I)) = Q(D)$ holds for all D – the modification strategy maintains consistency with invariants.

The inputs to the modification strategy, $D_{-\mathcal{S}}$ and I, contain no information for

the respondents in $\mathcal{S}$, except what is necessary to compute the invariants I. As in the single record modification strategy case, we can consider two scenarios:

1. *Actual Release*: $Q(D) = I, M(I, D) = \omega$.
2. *Private Counterfactual*: $\varphi_{\mathcal{S}}$ is used everywhere, $Q(\varphi_{\mathcal{S}}(D_{-\mathcal{S}}, I)) = I, M(I, \varphi_{\mathcal{S}}(D_{-\mathcal{S}}, I)) = \omega$.

In Examples 3 and 4 there is no single record modification strategy that can be interpreted as private because the only record that can be added to D_{-A} is the true record r_A (any one-record change to the database invalidates the invariants). It is, however, possible to provide some form of guarantee using a group modification strategy. In the case of Examples 3 and 4, the set $\mathcal{S}$ would include all of the respondents.

The following theorem provides posterior-to-posterior semantics that relate the privacy provided by the private counterfactual world to the privacy provided when a mechanism M satisfying ϵ-differential privacy is used in the actual world.

Theorem 6 Let D be a random variable distributed according to the joint distribution of all the records in a database, $\pi(D = d)$ be an arbitrary attacker's prior supported over $d \in X^n$ about all records in the database. Let Q be a deterministic algorithm for computing invariants, M be a bounded ϵ-differentially private mechanism that takes both invariants and a database as input. Let $\mathcal{S} = \{A_1, A_2, \ldots, A_k\}$ be a set of k respondents from the database, $\mathbf{R}$ represent the joint distribution of

any subset of records from respondents $A_1, \ldots, A_k$ and $\vec{r}$ be a possible record vector value. Let $\varphi_{\mathcal{S}}$ be a group modification strategy for $\mathcal{S}$. Then for fixed I and any ω,

$$\frac{\pi(\mathbf{R} = \vec{r} \mid \mathrm{Q}(D) = I, M(I, D) = \omega)}{\pi(\mathbf{R} = \vec{r} \mid \mathrm{Q}(\varphi_{\mathcal{S}}(D_{-\mathcal{S}}, I)) = I, M(I, \varphi_{\mathcal{S}}(D_{-\mathcal{S}}, I)) = \omega)} \in \left[e^{-2|\mathcal{S}|\epsilon}, e^{2|\mathcal{S}|\epsilon}\right]$$

See Appendix C for the proof. This leads to a similar interpretation of privacy guarantees for respondents in $\mathcal{S}$ (collectively or individually) to that of Theorem 5. The inference an arbitrary attacker can make about any subsets of respondents in $\mathcal{S}$ is at most $e^{2|\mathcal{S}|\epsilon}$ times sharper than in the private counterfactual in which a group modification is used. In the private counterfactual, only invariant statistics aggregated for the entire group are used by the mechanism. Therefore no information regarding the specific assignment of invariant-related variables to specific group members or any variables not related to the computation of invariants for the group member are utilized by the mechanism in the private counterfactual. Assuming that the set $\mathcal{S}$ has been chosen such that there are multiple possible different assignments of attributes to group members (i.e., uncertainty), then the counterfactual may be considered private for the members of $\mathcal{S}$ (acknowledging the initial privacy loss due to the invariants). As before, the privacy loss due to the invariants themselves is generally not quantifiable. This establishes the two components of the privacy guarantee: (1) the counterfactual is private assuming the invariants are not themselves too disclosive and (2) for small ϵ and $|\mathcal{S}|$, attacker inference will be similar in the actual release and in the counterfactual.

In addition to the group privacy guarantee conditional on the invariants, the

group modification strategy also offers a more nuanced protection of invariant attributes for a respondent than the single record modification strategy. Recall that in the private counterfactual associated with the single record modification strategy, Respondent A still had to provide information about her invariant attributes in order to calculate I. By considering the differences between the invariant totals in D_{-A} and I, it would be possible to recover Respondent A's invariant attributes. As a result, no statement could be made about the protection of Respondent A's invariant attributes in the counterfactual world. However with a group modification strategy, Respondent A's invariant attributes are only provided as an aggregate along with the invariant attributes of the other respondents in $\mathcal{S}$. In this way, Respondent A's invariant attributes can be 'hidden' amongst the groups' aggregate invariant attributes, assuming some heterogeneity exists among the group's members invariant attributes. In order for this to be meaningful, Respondent A's invariant attributes must not be recoverable from $D_{-\mathcal{S}}$ and I. This protection offered to a respondent's invariant attributes in the counterfactual is different from that offered to non-invariant attributes in the sense that the invariant attribute is still used in the mechanism. However, since it can be hidden amongst the aggregate invariants of the group, the inference of some attackers in the counterfactual about Respondent A's invariant attributes would be weakened because of additional uncertainty, and therefore the counterfactual is in a sense more private for these attributes than with single record modification strategies. However, with the group modification strategy the posterior-posterior bound relative to the actual release is larger ($e^{2|\mathcal{S}|\epsilon}$ vs. $e^{2\epsilon}$).

3.9 Tools for Analyzing Leakage Guarantees as Odds Ratios

In this section we extend our privacy semantics with the use of group modification strategies to study the ability of an attacker to discriminate between alternatives.

For each person, we specify a set of secret pairs $\mathbb{S}_{\text{pairs}} = \{(\sigma_{a_1}, \sigma_{b_1}), \ldots, (\sigma_{a_\ell}, \sigma_{b_\ell})\}$. We are interested in protecting against an attacker using the output of M to improve his inference about how likely it is that σ_{a_i} is true about a person's record r compared to σ_{b_i} being true. We express those statements mathematically as either $\sigma_{a_i}(r) = \text{True}$ or $\sigma_{b_i}(r) = \text{True}$, because for each i, σ_{a_i} and σ_{b_i} are mutually exclusive.

To measure the relative likelihood of one secret compared to another, we treat the record as a random variable R (since the attacker does not know it) and we calculate the *odds*. The *odds* are the probability that $\sigma_{a_i}(R) =$True given an attacker's prior π, the output of M, and the output of Q divided by the probability that $\sigma_{b_i}(R) =$True given an attacker's prior π, the output of M, and the output of Q:

$$\frac{\pi(\sigma_{a_i}(R) = \text{True} \mid \text{Q}(D) = I, M(I, D) = \omega)}{\pi(\sigma_{b_i}(R) = \text{True} \mid \text{Q}(D) = I, M(I, D) = \omega)}$$

We compare these odds to the odds assuming a private counterfactual in which there is a group $\mathcal{S}$ of people whose information is removed and who only report the group statistics that are necessary to maintain the invariants. We employ a group modification strategy that replaces the deleted records with values that sat-

isfy the invariants. A properly chosen group will mean that there are several possible assignments of attributes to the record of Respondent A and in the private counterfactual no information about any of these possible assignments are reported. In this private counterfactual world, the odds are:

$$\frac{\pi(\sigma_{a_i}(R) = \text{True} \mid \text{Q}(D) = I, M(I, \varphi_{\mathcal{S}}(D_{-\mathcal{S}}, I)) = \omega)}{\pi(\sigma_{b_i}(R) = \text{True} \mid \text{Q}(D) = I, M(I, \varphi_{\mathcal{S}}(D_{-\mathcal{S}}, I)) = \omega)}$$

The privacy leakage is then the following odds ratio and represents the improvement in inference due to using true records rather than the modification strategy which scrubs records:

$$\begin{aligned}
&\frac{\dfrac{\pi(\sigma_{a_i}(R) = \text{True} \mid \text{Q}(D) = I, M(I, D) = \omega)}{\pi(\sigma_{b_i}(R) = \text{True} \mid \text{Q}(D) = I, M(I, D) = \omega)}}{\dfrac{\pi(\sigma_{a_i}(R) = \text{True} \mid \text{Q}(D) = I, M(I, \varphi_{\mathcal{S}}(D_{-\mathcal{S}}, I)) = \omega)}{\pi(\sigma_{b_i}(R) = \text{True} \mid \text{Q}(D) = I, M(I, \varphi_{\mathcal{S}}(D_{-\mathcal{S}}, I)) = \omega)}} \\
&= \frac{\dfrac{\pi(\sigma_{a_i}(R) = \text{True}, \text{Q}(D) = I, M(I, D) = \omega)}{\pi(\sigma_{b_i}(R) = \text{True}, \text{Q}(D) = I, M(I, D) = \omega)}}{\dfrac{\pi(\sigma_{a_i}(R) = \text{True}, \text{Q}(D) = I, M(I, \varphi_{\mathcal{S}}(D_{-\mathcal{S}}, I)) = \omega)}{\pi(\sigma_{b_i}(R) = \text{True}, \text{Q}(D) = I, M(I, \varphi_{\mathcal{S}}(D_{-\mathcal{S}}, I)) = \omega)}} \\
&= \frac{\dfrac{\sum_d \pi(\sigma_{a_i}(R) = \text{True}, \text{Q}(D) = I, D = d) P(M(I, d) = \omega)}{\sum_d \pi(\sigma_{b_i}(R) = \text{True}, \text{Q}(D) = I, D = d) P(M(I, d) = \omega)}}{\dfrac{\sum_d \pi(\sigma_{a_i}(R) = \text{True}, \text{Q}(D) = I, D = d) P(M(I, \varphi_{\mathcal{S}}(d_{-\mathcal{S}}, I)) = \omega)}{\sum_d \pi(\sigma_{b_i}(R) = \text{True}, \text{Q}(D) = I, D = d) P(M(I, \varphi_{\mathcal{S}}(d_{-\mathcal{S}}, I)) = \omega)}} \\
&\in [e^{-2\epsilon|\mathcal{S}|}, e^{2\epsilon|\mathcal{S}|}]
\end{aligned}$$

when M satisfies bounded ϵ-differential privacy. This result is summarized as follows:

Theorem 7 Let D be a random variable representing the true database, π be an arbitrary prior on D held by an attacker, Q be an algorithm for computing invariants, M be an bounded ϵ-differentially private mechanism that takes both invariants and a database as input. Let R be the random variable (in the view of the attacker) corresponding to the record for Respondent A. Let $\mathcal{S}$ be a group of people and let $\varphi_{\mathcal{S}}$ be a group modification strategy. Then for any fixed I and ω and any secret pair $(\sigma_{a_i}, \sigma_{b_i})$,

$$\frac{\dfrac{\pi(\sigma_{a_i}(R) = \text{True} \mid \mathrm{Q}(D) = I, M(I, D) = \omega)}{\pi(\sigma_{b_i}(R) = \text{True} \mid \mathrm{Q}(D) = I, M(I, D) = \omega)}}{\dfrac{\pi(\sigma_{a_i}(R) = \text{True} \mid \mathrm{Q}(D) = I, M(I, \varphi_S(D_{-\mathcal{S}}, I)) = \omega)}{\pi(\sigma_{b_i}(R) = \text{True} \mid \mathrm{Q}(D) = I, M(I, \varphi_S(D_{-\mathcal{S}}, I)) = \omega)}} \in [e^{-2\epsilon|\mathcal{S}|}, e^{2\epsilon|\mathcal{S}|}]$$

Note that Theorem 6 can be used directly to provide odds ratio bounds as well, but those bounds would be worse (i.e., $e^{4|\mathcal{S}|}$ instead of $e^{2|\mathcal{S}|}$).

3.10 Guarantees for Attributes Involved in Computation of Census Invariants

In Section 3.7, we considered the privacy impact of invariants under Theorem 5. In short, attributes involved in the computation of invariants were not given quantifiable protection but the rest of the attributes had quantifiable protections. In this section, we use Theorem 6 and Theorem 7 to provide more nuanced guarantees for these attributes. The schema and invariants are as defined in Section 3.7.

Case 1: **C1** only.

Per Theorem 7, the inference an attacker can make about a respondent's block after the mechanism's release is at best $e^{4\epsilon}$ times better than if the respondent's block was swapped with the block of another respondent.

Case 2: **C1, C2**

Per Theorem 7, the inference an attacker can make about a respondent's block and voting-age status after the mechanism's release is at best $e^{8\epsilon}$ times better than if the voting-age status and block of the respondent and 3 other individuals were arbitrarily reassigned. More specifically, for any group of 4 individuals consisting of one voting age and one non-voting age person in one block and one voting age and one non-voting age person in a second block, the assignment of block and voting-age status to each individual is protected by an inference bound of $e^{8\epsilon}$ compared to a private counterfactual in which those attributes in their records were randomly re-assigned among the 4 individuals.

Case 3: **C1, C3**

Under the basic analysis of Section 3.7, there were no quantifiable protections of housing status and block, and in some cases age and sex. However, with Theorem 7, the inference an attacker can make about a respondent's record is at most $e^{4\epsilon}$ times better than if the block and housing status of a respondent were swapped

with that of another individual, and the rest of the record was altered arbitrarily.

Case 4: **C1, C4**

The privacy guarantees provided by Theorem 7 are the same as in Case 3.

Case 5: **C1, C5**

Under Theorem 7, the inference an attacker can make about a respondent's record is at most $e^{4\epsilon}$ times better than if the block and housing status of a respondent were swapped with that of another individual, and the other values in this pair of records were modified arbitrarily subject to edit rules. For example, if the respondent is from a block containing only female-only dormitories, then after swapping blocks with another individual, that individual's reported sex would be changed to female by edit rules. However, the assignment of which individual belongs to which block and has which sex is protected with an inference bound of at most $e^{4\epsilon}$ compared to a random re-assignment of blocks to the pair of individuals.

Case 6: **C3, C5**

Vacant group quarters are not tabulated, and so the invariants reveal a minimum on the population in each block, and Theorem 7 is needed. For individuals

living in group quarters containing only one person, as per Theorem 7, inference about whether housing status is protected by an inference bound of at most $e^{4\epsilon}$ than if the housing status of the respondent's record was swapped with that of another record and both records were modified arbitrarily (subject to edit rules as in case 5). For individuals not living in a single-person group quarters, as per Theorem 7, inference about any part of the response is protected by an inference bound of at most $e^{2\epsilon}$ than if the entire record was arbitrarily modified.

Case 7: **C1, C3, C5**

The analysis is similar to Cases 3 and 5. Under Theorem 7, the inference an attacker can make about a respondent's record is at most $e^{4\epsilon}$ times better than if the block of a respondent was swapped with the block of another individual, and the other values in this pair of records were modified arbitrarily subject to edit rules (as in Case 5).

Case 8: **C1, C2, C3, C4, C5**

When all five invariants are used, the privacy guarantees are most complicated and have the weakest upper bound in inference. If the respondent is in any group of 8 people, with 4 from one block (having all combinations of voting age status and housing status) and 4 from another block (having all combinations of voting status and housing status), then Theorem 7 guarantees that the inference an at-

tacker can make about a respondent's record is at most $e^{16\epsilon}$ times better than if the assignment of block, voting age status and housing status were randomly reassigned within the group.

3.11 Conclusions

The release of exact statistics (i.e., invariants) about a dataset causes a privacy leakage amplification where the normal privacy leakage of a disclosure control algorithm gets amplified when its output is combined with invariants. In this paper, we extended the semantics of differential privacy to analyze this amplified privacy leakage. The analysis required the use of group modification strategies which are designed to remove information before any algorithm computes on the data. The resulting privacy guarantees relate the posterior inference about an individual in actuality compared to the private counterfactual world in which the modification strategies were used.

We applied these results to the invariants under consideration for the 2020 Decennial Census of Population and Housing. The information leakage due to the invariants is generally not quantifiable, but any subsequent leakage (e.g., when the invariants are combined with the output of a differentially private algorithm) can be quantified and separated into leakage due to disclosure control and leakage due to the amplification caused by the invariants.

APPENDIX A

APPENDIX OF CHAPTER 1

A.0.1 Formal definition of neighboring databases

We can represent any database D by its un-normalized histogram $x \in \mathbb{Z}^{*|\chi|}$. The notation $|\chi|$ represents the cardinality of the set χ, from which database entries (rows) are drawn, and $\mathbb{Z}^*$ is the set of non-negative integers. Each entry in x, x_i, is the number of elements in the database D of type $i \in \chi$. The ℓ_1 norm of x is

$$||x||_1 = \sum_{i=1}^{|\chi|} |x_i| \,. \tag{A.1}$$

Observe that $||x||_1 = N$, 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. We define *adjacent histograms* as those with equal ℓ_1 norm and between which the ℓ_1 distance is 2.

If x is the histogram representation of D, y is the histogram representation of D', and D' is constructed from D by modifying exactly one row, then $||x||_1 = ||y||_1$ and $||x-y||_1 = 2$. So, D and D' are adjacent databases and x and y are the adjacent histogram representations of D and D', respectively. Some caution is required when reviewing the related literature because definitions may be stated in terms of adjacent databases or adjacent histograms.

A.0.2 Translation of the Ghosh-Roth Model to Our Notation

In this appendix we show that the results in our Section 1.3, based on the definitions in the text using database histograms and normalized queries, are equivalent to the results in Ghosh and Roth (2015). In what follows, definitions and theorems tagged GR refer to the original Ghosh and Roth (GR, hereafter) paper. Untagged definitions and theorems refer to our results in the text.

GR model a database $D \in \{0,1\}^n$ where there is a single bit, b_i, taking values in $\{0,1\}$ for a population of individuals $i = 1, \ldots, n$. In GR-Definition 2.1, they define a query release mechanism $A(D)$, a randomized algorithm that maps $\{0,1\}^n \to \mathbb{R}$, as ε_i-differentially private if for all measurable subsets S of $\mathbb{R}$ and for any pair of databases D and $D^{(i)}$ such that $H\left(D, D^{(i)}\right) = 1$

$$\frac{\Pr\left[A(D) \in S\right]}{\Pr\left[A\left(D^{(i)}\right) \in S\right]} \leqslant e^{\varepsilon_i}$$

where $H\left(D, D^{(i)}\right)$ is the Hamming distance between D and $D^{(i)}$. Notice that their use of the Hamming distance to define neighboring databases is consistent with our use of "bounded" differential privacy. However, this is not the standard definition of ε-differential privacy, which they take from Dwork et al. (2006), because a "worst-case" extremum is not included. The parameter ε_i is specific to individual i. The amount of privacy loss algorithm A permits for individual i, whose bit b_i is the one that is toggled in $D^{(i)}$, is potentially different from the privacy loss allowed for individual $j \neq i$, whose privacy loss may be $\varepsilon_j > \varepsilon_i$ from the same algorithm. In this case, individual j could also achieve ε_j-differentially privacy if

the parameter ε_i were substituted for ε_j. To refine this definition so that it also corresponds to an extremum with respect to each individual, GR-Definition 2.1 adds the condition that algorithm A is ε_i-*minimally differentially private* with respect to individual i if

$$\varepsilon_i = \arg\inf_{\varepsilon} \left\{ \frac{\Pr\left[A\left(D\right) \in S\right]}{\Pr\left[A\left(D^{(i)}\right) \in S\right]} \leqslant e^{\varepsilon} \right\},$$

which means that for individual i, the level of differential privacy afforded by the algorithm $A\left(D\right)$ is the smallest value of ε for which algorithm A achieves ε-differential privacy for individual i. In GR ε_i-differentially private always means ε_i-minimally differentially private.

GR-Fact 1, stated without proof, but see Dwork and Roth (2014, p. 42-43) for a proof, says that ε_i-minimal differential privacy composes. That is, if algorithm $A\left(D\right)$ is ε_i-minimally differentially private, $T \subset \{1, \ldots, n\}$, and $D, D^{(T)} \in \{0,1\}^n$ with $H\left(D, D^{(T)}\right) = |T|$, then

$$\frac{\Pr\left[A\left(D\right) \in S\right]}{\Pr\left[A\left(D^{(T)}\right) \in S\right]} \leqslant e^{\left\{\sum_{i \in T} \varepsilon_i\right\}},$$

where $D^{(T)}$ differs from D only on the indices in T.

In the population, the statistic of interest is an unnormalized query

$$s = \sum_{i=1}^{n} b_i.$$

The ε_i-minimally differentially private algorithm $A\left(D\right)$ delivers an output $\hat{s}$ that is a noisy estimate of s, where the noise is induced by randomness in the query release mechanism embedded in A. Each individual in the population when offered a payment $p_i > 0$ in exchange for the privacy loss $\varepsilon_i > 0$ computes an

individual privacy cost equal to $\upsilon_i \varepsilon_i$, where $\upsilon_i > 0$, $p \equiv (p_1, \ldots, p_n) \in \mathbb{R}_+^n$, and $\upsilon \equiv (\upsilon_1, \ldots, \upsilon_n) \in \mathbb{R}_+^n$.

GR define a mechanism M as a function that maps $\mathbb{R}_+^n \times \{0,1\}^n \to \mathbb{R} \times \mathbb{R}_+^n$ using an algorithm $A(D)$ that is $\varepsilon_i(\upsilon)$-minimally differentially private to deliver a query response $\hat{s} \in \mathbb{R}$ and a vector of payments $p(\upsilon) \in \mathbb{R}_+^n$. GR-Definition 2.4 defines individually rational mechanisms. GR-Definition 2.5 defines dominant-strategy truthful mechanisms. An individually rational, dominant-strategy truthful mechanism M provides individual i with utility $p_i(\upsilon) - \upsilon_i \varepsilon_i(\upsilon) \geqslant 0$ and $p_i(\upsilon) - \upsilon_i \varepsilon_i(\upsilon) \geqslant p_i\left(\upsilon^{\tilde{}i}, \upsilon_i'\right) - \upsilon_i \varepsilon_i\left(\upsilon^{\tilde{}i}, \upsilon_i'\right)$ for all $\upsilon_i' \in \mathbb{R}_+^n$, where $\upsilon^{\tilde{}i}$ is the vector υ with element υ_i removed.

GR define k-accuracy in GR-Definition 2.6 using the deviation $|\hat{s} - s|$ from the output $\hat{s}$ produced by algorithm $A(D)$ using mechanism M as

$$\Pr\left[|\hat{s} - s| \geqslant k\right] \leqslant \frac{1}{3}.$$

where we have reversed the direction of the inequalities and taken the complementary probability to show that this is the unnormalized version of our Definition 3 for a query sequence of length 1. GR also define the normalized query accuracy level as α, which is identical to our usage in Definition 3.

GR-Theorem 3.1 uses the GR definitions of ε_i-minimal differential privacy, k-accuracy, and GR-Fact 1 composition to establish that any differentially private mechanism M that is $\left(\frac{\alpha n}{4}\right)$-accurate must purchase privacy loss of at least $\varepsilon_i \geqslant \frac{1}{\alpha n}$ from at least $H \geqslant (1-\alpha)n$ individuals in the population. GR-Theorem 3.3

establishes the existence of a differentially private mechanism that is $\left(\frac{1}{2} + \ln 3\right) \alpha n$-accurate and selects a set of individuals $H \subset \{1, \ldots, n\}$ with $\varepsilon_i = \frac{1}{\alpha n}$ for all $i \in H$ and $|H| = (1 - \alpha)\, n$.

In order to understand the implications of GR-Theorems 3.1 and 3.3 and our arguments about the public-good properties of differential privacy, consider the application of GR-Definition 2.3 ($\mathrm{Lap}\,(\sigma)$ noise addition) to construct an ε-differentially private response to the counting query based on GR-Theorem 3.3 with $|H| = (1 - \alpha)\, n$ and the indices ordered such that $H = \{1, \ldots, |H|\}$. The resulting answer from the query response mechanism is

$$\hat{s} = \sum_{i=1}^{H} b_i + \frac{\alpha n}{2} + \mathrm{Lap}\left(\frac{1}{\varepsilon}\right),$$

which is the counting query version of equation (1.3) in the text. Note the bias correction term $\alpha n/2$ is adjusted in equation (1.3) as necessitated by our use of $\left(\alpha, \frac{1}{3}\right)$-accuracy. Because of GR-Theorem 3.3, we can use a common $\varepsilon = \frac{1}{\alpha n}$ in equation (1.3).

If this were not true, then we would have to consider query release mechanisms that had different values of ε for each individual in the population and therefore the value that enters equation (1.3) would be much more complicated. To ensure that each individual in H received ε_i-minimally differential privacy, the algorithm would have to use the smallest ε_i that was produced for any individual. In addition, the FairQuery and MinCostAuction algorithms described next would not work because they depend upon being able to order the cost functions $\upsilon_i \varepsilon_i$ by υ_i, which is not possible unless ε_i is a constant or υ_i and ε_i are perfectly

positively correlated. Effectively, GR-Theorem 3.3 proves that achieving (α, β)-accuracy with ε-differential privacy requires a mechanism in which everyone who sells a data-use right gets the best protection (minimum ε_i over all $i \in H$) offered to anyone in the analysis sample. If a change in the algorithm's parameters results in a lower minimum ε_i, everyone who opts to use the new parameterization receives this improvement. In addition, we argue in the text that when such mechanisms are used by a government agency they are also non-excludable because exclusion from the database violates equal protection provisions of the laws that govern these agencies.

Next, GR analyze algorithms that achieve $O(an)$-accuracy by purchasing exactly $\frac{1}{\alpha n}$ units of privacy loss from exactly $(1-\alpha)n$ individuals. Their algorithms *FairQuery* and *MinCostAuction* have the same basic structure:

- Sort the individuals in increasing order of their privacy cost, $\upsilon_1 \leqslant \upsilon_2 \leqslant \ldots \leqslant \upsilon_n$.
- Find the cut-off value υ_k that either exhausts a budget constraint (FairQuery) or meets an accuracy constraint (MinCostAuction).
- Assign the set $H = \{1, \ldots, k\}$.
- Calculate the statistic $\hat{s}$ using a differentially private algorithm that adds Laplace noise with just enough dispersion to achieve the required differential privacy for the privacy loss purchased from the members of H.
- Pay all members of H the same amount, a function of υ_{k+1}; pay all others nothing.

To complete the summary of GR, we note that GR-Theorem 4.1 establishes that FairQuery is dominant-strategy truthful and individually rational. GR-Proposition 4.4 establishes that FairQuery maximizes accuracy for a given total

privacy purchase budget in the class of all dominant-strategy truthful, individually rational, envy-free, fixed-purchase mechanisms. GR-Proposition 4.5 proves that their algorithm MinCostAuction is a VCG mechanism that is dominant-strategy truthful, individually rational and $O(\alpha n)$-accurate. GR-Theorem 4.6 provides a lower bound on the total cost of purchasing k units of privacy of kv_{k+1} GR-Theorem 5.1 establishes that for $v \in \mathbb{R}_+^n$, no individually rational mechanism can protect the privacy of valuations v with (k, β)-accuracy for $k < \frac{n}{2}$.

In our application of GR, we use N as the total population. Our γ_i is identical to the GR v_i. We define the query as a normalized query, which means that query accuracy is defined in terms of α instead of k; hence, our implementation of the VCG mechanism achieves (α, β) where the inclusion of β generalizes GR's implicit restriction to $\beta = \frac{1}{3}$ in their accuracy definition. We define the individual amount of privacy loss in the same manner as GR.

APPENDIX B

APPENDIX OF CHAPTER 2

Suppose a Bayesian adversary wants to learn the record R belonging to individual i, from a confidential database, x. She has auxiliary information E that includes traditional identifiers (e.g., name and address) along with other variables that can be used to match against data published via differential privacy. The adversary has prior μ over the space of possible data vectors $\mathcal{D}$. A data custodian uses a bounded ϵ-differentially private mechanism M to publish output $M(x) = \omega$. Bounded differential privacy mechanisms treat the total number of records in the confidential database as public. Unbounded differential privacy mechanisms inject noise into the total record count as well. The algorithms under consideration for use with the 2020 Census are in the class of bounded differential privacy mechanisms. Upon observing ω and E, the adversary updates her beliefs about R, the record of an individual i, using Bayes law. By the law of total probability,

$$\mu(R = r|\omega, E) = \sum_{z \in \mathcal{D}} \mu(R = r, z|\omega, E)$$

Note that

$$\begin{aligned}
\mu(R = r, z|\omega, E) &= \frac{\mu(R = r, \omega, E|z)\mu(z)}{\mu(\omega, E)} \\
&= \frac{\mu(R = r, E|z)Pr[M(z) = \omega]\mu(z)}{\sum_{y \in \mathcal{D}} \mu(\omega, E|y)\mu(y)} \\
&= \frac{\mu(R = r, E|z)Pr[M(z) = \omega]\mu(z)}{\sum_{y \in \mathcal{D}} \mu(E|y)Pr[M(y) = \omega]\mu(y)},
\end{aligned}$$

where the second equality follows under the assumption that ω is conditionally independent from R and E given z. The probability of observing ω given z is completely determined by the coin flips of the mechanism. Hence,

$$\mu(R = r|\omega, E) = \frac{\sum_{z \in \mathcal{D}} \mu(R = r, E, z) Pr[M(z) = \omega]}{\sum_{y \in \mathcal{D}} \mu(E, y) Pr[M(y) = \omega]}.$$

Now consider a hypothetical counterfactual where the mechanism M does not use i's record, and the adversary knows it. Instead M runs on $\tilde{x} = x_{-i} \cup r_f$ the data vector in which i's record is removed from x and replaced by an arbitrary default record, r_f. In this case, the adversary's updated beliefs are:

$$\mu_{-i}(R = r|\omega, E) = \frac{\sum_{z \in \mathcal{D}} \mu(R = r, E, z) Pr[M(\tilde{z}) = \omega]}{\sum_{y \in \mathcal{D}} \mu(E, y) Pr[M(\tilde{y}) = \omega]}.$$

The notation μ_{-i} characterizes beliefs over $\tilde{x}$ derived from μ and knowledge that

R has been removed and replaced by r_f. We conclude the following:

$$
\begin{aligned}
&\frac{\mu(R=r|\omega,E)}{\mu_{-i}(R=r|\omega,E)} \\
&= \frac{\sum_{z\in\mathcal{D}}\mu(R=r,E,z)Pr[M(z)=\omega]/\sum_{y\in\mathcal{D}}\mu(E,y)Pr[M(y)=\omega]}{\sum_{z\in\mathcal{D}}\mu(R=r,E,z)Pr[M(\tilde{z})=\omega]/\sum_{y\in\mathcal{D}}\mu(E,y)Pr[M(\tilde{y})=\omega]} \\
&= \frac{\sum_{z\in\mathcal{D}}\mu(R=r,E,z)Pr[M(z)=\omega]/\sum_{z\in\mathcal{D}}\mu(R=r,E,z)Pr[M(\tilde{z})=\omega]}{\sum_{y\in\mathcal{D}}\mu(E,y)Pr[M(y)=\omega]/\sum_{y\in\mathcal{D}}\mu(E,y)Pr[M(\tilde{y})=\omega]} \\
&\leqslant \frac{\sum_{z\in\mathcal{D}}\mu(R=r,E,z)e^{\epsilon}Pr[M(\tilde{z})=\omega]/\sum_{z\in\mathcal{D}}\mu(R=r,E,z)Pr[M(\tilde{z})=\omega]}{\sum_{y\in\mathcal{D}}\mu(E,y)Pr[M(y)=\omega]/\sum_{y\in\mathcal{D}}\mu(E,y)Pr[M(\tilde{y})=\omega]} \\
&\quad (M \text{ is bounded } \epsilon\text{-differentially private so } \frac{Pr[M(z)=\omega]}{Pr[M(\tilde{z})=\omega]} \leqslant e^{\epsilon}.) \\
&= \frac{e^{\epsilon}\sum_{z\in\mathcal{D}}\mu(R=r,E,z)Pr[M(\tilde{z})=\omega]/\sum_{z\in\mathcal{D}}\mu(R=r,E,z)Pr[M(\tilde{z})=\omega]}{\sum_{y\in\mathcal{D}}\mu(E,y)Pr[M(y)=\omega]/\sum_{y\in\mathcal{D}}\mu(E,y)Pr[M(\tilde{y})=\omega]} \\
&\quad \text{(Factor out } e^{\epsilon}.) \\
&= \frac{e^{\epsilon}}{\sum_{y\in\mathcal{D}}\mu(E,y)Pr[M(y)=\omega]/\sum_{y\in\mathcal{D}}\mu(E,y)Pr[M(\tilde{y})=\omega]} \\
&\quad \text{(The summations in the numerator ratio cancel out.)} \\
&\leqslant \frac{e^{\epsilon}}{\sum_{y\in\mathcal{D}}\mu(E,y)e^{-\epsilon}Pr[M(\tilde{y})=\omega]/\sum_{y\in\mathcal{D}}\mu(E,y)Pr[M(\tilde{y})=\omega]} \\
&\quad (M \text{ is bounded } \epsilon\text{-differentially private so } \frac{Pr[M(y)=\omega]}{Pr[M(\tilde{y})=\omega]} \geqslant e^{-\epsilon}.) \\
&= \frac{e^{\epsilon}}{e^{-\epsilon}\sum_{y\in\mathcal{D}}\mu(E,y)Pr[M(\tilde{y})=\omega]/\sum_{y\in\mathcal{D}}\mu(E,y)Pr[M(\tilde{y})=\omega]} \\
&\quad \text{(Factor out } e^{-\epsilon}) \\
&= e^{2\epsilon} \\
&\quad \text{(The summations in the denominator ratio cancel out.)}
\end{aligned}
$$

Similarly, $\frac{\mu(R=r|\omega,E)}{\mu_{-i}(R=r|\omega,E)} \geqslant e^{-2\epsilon}$.

APPENDIX C

APPENDIX OF CHAPTER 3

Proof of Equation 3.2 **Proof.**

$$\pi(R_A = r_1 \mid M(D) = \omega) = \frac{\pi(R_A = r_1, M(D) = \omega)}{Pr(M(D) = \omega)}$$

and similarly for $\pi(R_A = r_2 \mid M(D) = \omega)$, $\pi(R_A = r_1 \mid M(D') = \omega)$, and $\pi(R_A = r_2 \mid M(D') = \omega)$. Therefore

$$\begin{aligned}
&\frac{\pi(R_A = r_1 \mid M(D) = \omega) / \pi(R_A = r_2 \mid M(D) = \omega)}{\pi(R_A = r_1 \mid M(D') = \omega) / \pi(R_A = r_2 \mid M(D') = \omega)} \\
&= \frac{\pi(R_A = r_1, M(D) = \omega) / \pi(R_A = r_2, M(D) = \omega)}{\pi(R_A = r_1, M(D') = \omega) / \pi(R_A = r_2, M(D') = \omega)} \\
&= \frac{\sum\limits_{\substack{d \in X^n \\ d\, s.t. R_A = r_1}} \pi(D = d) Pr(M(d) = \omega) \Big/ \sum\limits_{\substack{d \in X^n \\ d\, s.t. R_A = r_2}} \pi(D = d) Pr(M(d) = \omega)}{\sum\limits_{\substack{d \in X^n \\ d\, s.t. R_A = r_1}} \pi(D = d) Pr(M(d') = \omega) \Big/ \sum\limits_{\substack{d \in X^n \\ d\, s.t. R_A = r_2}} \pi(D = d) Pr(M(d') = \omega)} \\
&\in \left[e^{-2\epsilon}, e^{2\epsilon}\right]
\end{aligned}$$

because

$$\frac{\sum\limits_{\substack{d \in X^n \\ d\, s.t. R_A = r_1}} \pi(D = d) Pr(M(d) = \omega)}{\sum\limits_{\substack{d \in X^n \\ d\, s.t. R_A = r_1}} \pi(D = d) Pr(M(d') = \omega)} \in \left[e^{-\epsilon}, e^{\epsilon}\right]$$

and

$$\frac{\sum\limits_{\substack{d \in X^n \\ d\, s.t. R_A = r_2}} \pi(D = d) Pr(M(d') = \omega)}{\sum\limits_{\substack{d \in X^n \\ d\, s.t. R_A = r_2}} \pi(D = d) Pr(M(d) = \omega)} \in \left[e^{-\epsilon}, e^{\epsilon}\right].$$

■

Proof of Theorem 5 **Proof.** Let $\mathbf{1}()$ be the indicator function and denote the dataset after the modification strategy as $D' = \phi_A(D_{-A}, I)$ or as $d' = \phi_A(d_{-A}, I)$ for dataset realization d.

$$
\begin{aligned}
&\frac{\pi(R_A = r \mid \mathrm{Q}(D) = I, M(I, D) = \omega)}{\pi(R_A = r \mid \mathrm{Q}(\phi_A(D_{-A}, I)) = I, M(I, \phi_A(D_{-A}, I)) = \omega)} \\
&\quad = \frac{\pi(R_A = r \mid \mathrm{Q}(D) = I, M(I, D) = \omega)}{\pi(R_A = r \mid \mathrm{Q}(D') = I, M(I, D') = \omega)} \\
&\quad = \frac{\dfrac{\sum_{d \in X^n} \pi(R_A = r, \mathrm{Q}(d) = I, M(I, d) = \omega, D = d)}{\sum_{d \in X^n} \pi(\mathrm{Q}(d) = I, M(I, d) = \omega, D = d)}}{\dfrac{\sum_{d \in X^n} \pi(R_A = r, \mathrm{Q}(d' = I, M(I, d') = \omega, D = d)}{\sum_{d \in X^n} \pi(\mathrm{Q}(d') = I, M(I, d') = \omega, D = d)}} \\
&\quad = \frac{\dfrac{\sum_{d \in X^n} \pi(R_A = r, D = d)\mathbf{1}(\mathrm{Q}(d) = I) Pr(M(I, d) = \omega)}{\sum_{d \in X^n} Pr(M(I, d) = \omega)\mathbf{1}(\mathrm{Q}(d) = I)\pi(D = d)}}{\dfrac{\sum_{d \in X^n} \pi(R_A = r, D = d)\mathbf{1}(\mathrm{Q}(d') = I) Pr(M(I, d' = \omega)}{\sum_{d \in X^n} Pr(M(I, d') = \omega)\mathbf{1}(\mathrm{Q}(d') = I)\pi(D = d)}} \\
&\quad \in [e^{-2\epsilon}, e^{2\epsilon}]
\end{aligned}
$$

because $\mathbf{1}(\mathrm{Q}(d) = I) = \mathbf{1}(\mathrm{Q}(d') = I)$ for all d as a result of the modification strategy definition and $Pr(M(I, d) = \omega)/P(M(I, d') = \omega) \in [e^{-\epsilon}, e^{\epsilon}]$ and d, d' differ by one record. This proof reflects the deterministic case. For randomized ϕ_A, condition on the randomness of ϕ_A until the end, and take the expectation with respect to the randomness in ϕ_A. ■

Proof of Theorem 6

Proof. Let $\mathbf{1}()$ be the indicator function and denote the dataset after the group modification strategy as $D' = \varphi_{\mathcal{S}}(D_{-\mathcal{S}}, I)$ or as $d' = \varphi_{\mathcal{S}}(d_{-\mathcal{S}}, I)$ for dataset realization d.

$$
\begin{aligned}
&\frac{\pi(\mathbf{R} = \vec{r} \mid \mathrm{Q}(D) = I, M(I, D) = \omega)}{\pi(\mathbf{R} = \vec{r} \mid \mathrm{Q}(\varphi_{\mathcal{S}}(D_{-\mathcal{S}}, I)) = I, M(I, \varphi_{\mathcal{S}}(D_{-\mathcal{S}}, I)) = \omega)} \\
&= \frac{\pi(\mathbf{R} = \vec{r} \mid \mathrm{Q}(D) = I, M(I, D) = \omega)}{\pi(\mathbf{R} = \vec{r} \mid \mathrm{Q}(D') = I, M(I, D') = \omega)} \\
&= \frac{\dfrac{\sum_{d \in X^n} \pi(\mathbf{R} = \vec{r}, \mathrm{Q}(d) = I, M(I, d) = \omega, D = d)}{\sum_{d \in X^n} \pi(\mathrm{Q}(d) = I, M(I, d) = \omega, D = d)}}{\dfrac{\sum_{d \in X^n} \pi(\mathbf{R} = \vec{r}, \mathrm{Q}(d') = I, M(I, d') = \omega, D = d)}{\sum_{d \in X^n} \pi(\mathrm{Q}(d') = I, M(I, d' = \omega, D = d)}} \\
&= \frac{\dfrac{\sum_{d \in X^n} \pi(\mathbf{R} = \vec{r}, D = d)\mathbf{1}(\mathrm{Q}(d) = I) Pr(M(I, d) = \omega)}{\sum_{d \in X^n} Pr(M(I, d) = \omega)\mathbf{1}(\mathrm{Q}(d) = I)\pi(D = d)}}{\dfrac{\sum_{d \in X^n} \pi(\mathbf{R} = \vec{r}, D = d)\mathbf{1}(\mathrm{Q}(d') = I) Pr(M(I, d' = \omega)}{\sum_{d \in X^n} Pr(M(I, d', I)) = \omega)\mathbf{1}(\mathrm{Q}(d) = I)\pi(D = d)}} \\
&\in \left[e^{-2|\mathcal{S}|\epsilon}, e^{2|\mathcal{S}|\epsilon}\right]
\end{aligned}
$$

because $\mathbf{1}(\mathrm{Q}(d) = I) = \mathbf{1}(\mathrm{Q}(d') = I)$ for all d as a result of the group modification strategy definition and $Pr(M(I, d) = \omega)/P(M(I, d') = \omega) \in [e^{-\epsilon|\mathcal{S}|}, e^{|\mathcal{S}|\epsilon}]$ as d and $d' = \varphi_{\mathcal{S}}(d_{-\mathcal{S}}, I)$ differ by at most $|\mathcal{S}|$ records. This proof reflects the deterministic case. For randomized $\varphi_{\mathcal{S}}$, condition on the randomness of $\varphi_{\mathcal{S}}$ until the end and

take the expectation with respect to the randomness in $\varphi_{\mathcal{S}}$. ■

BIBLIOGRAPHY

Abowd, J. M. (2019). Staring down the database reconstruction theorem, *American Association for the Advancement of Science Annual Meeting* .

Abowd, J. M. and Schmutte, I. M. (2015). Economic analysis and statistical disclosure limitation, *Brookings Papers on Economic Activity* pp. 221–267. Spring.

Abowd, J. M. and Schmutte, I. M. (2019). An economic analysis of privacy protection and statistical accuracy as social choices, *American Economic Review* **109**(1): 171–202.

Abowd, J. M., Schmutte, I. M., Sexton, W. N. and Vilhuber, L. (2019). Why the economics profession must actively participate in the privacy protection debate, *AEA Papers and Proceedings* **109**: 397–402.

Abowd, J. M., Stephens, B. E., Vilhuber, L., Andersson, F., McKinney, K. L., Roemer, M. and Woodcock, S. D. (2009). *The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators*, University of Chicago Press.

Acquisti, A., John, L. K. and Loewenstein, G. (2013). What is privacy worth?, *Journal of Legal Studies* **42**(2): 249–274.

Acquisti, A., Taylor, C. and Wagman, L. (2016). The economics of privacy, *Journal of Economic Literature* **54**(2): 442–492.

Antenucci, D., Cafarella, M., Levenstein, M., Re, C. and Shapiro, M. D. (2014). Using social media to measure labor market flows, *Working Paper 20010*, National Bureau of Economic Research.

Bassily, R., Groce, A., Katz, J. and Smith, A. D. (2013). Coupled-worlds privacy: Exploiting adversarial uncertainty in statistical data privacy, *FOCS*, pp. 439–448.

Benedetto, G., Stanley, J. C. and Totty, E. (2018). The creation and use of the sipp synthetic beta v7.0, *Technical report*, U.S. Census Bureau.

Brenner, H. and Nissim, K. (2014). Impossibility of differentially private universally optimal mechanisms, *SIAM Journal on Computing* **43**(5): 1513–1540.

Bureau, U. C. (2012). 2010 summary file 1 technical documentation, https://www.census.gov/prod/cen2010/doc/sf1.pdf.

Cavallo, A. and Rigobon, R. (2016). The billion prices project: Using online prices for measurement and research, *Journal of Economic Perspectives* **30**(2): 151–78.

Census DAS Team (2019a). Das 2010 demonstration data products release. https://github.com/uscensusbureau/census2020-das-2010ddp.

Census DAS Team (2019b). DAS E2E Release. https://github.com/uscensusbureau/census2020-das-e2e.

Chetty, R. and Friedman, J. (2019). A practical method to reduce privacy loss when disclosing statistics based on small samples, *Journal of Privacy and Confidentiality* **9**.

Chetty, R., Friedman, J., Hendren, N., Jones, M. and Porter, S. (2018). The opportunity atlas: Mapping the childhood roots of social mobility, *Working Paper w25147*, National Bureau of Economic Research.

Choi, H. and Varian, H. (2012). Predicting the present with Google trends, *Economic Record* **88**: 2–9.

Commission on Evidence-Based Policymaking (2017). The promise of evidence-based policymaking: Report of the Commission on Evidence-Based Policymaking, *Technical report*, Government Printing Office. https://www.cep.gov/content/dam/cep/report/cep-final-report.pdf.

Differential Privacy Team (2017). Learning with privacy at scale, *Apple Machine Learning Journal* **1**(8). https://machinelearning.apple.com/2017/12/06/learning-with-privacy-at-scale.html.

Ding, B., Kulkarni, J. and Yekhanin, S. (2017). Collecting telemetry data privately, *Advances in Neural Information Processing Systems 30* .

Dinur, I. and Nissim, K. (2003). Revealing information while preserving privacy, *PODS '03 Proceedings of the 22nd ACM SIGMOD-SIGACT-SIGART symposium on Principles of Database Systems* pp. 202–210.

Dwork, C. (2006). Differential privacy, *Proceedings of the 33rd International Colloquium on Automata, Languages and Programming* pp. 1–12.

Dwork, C., McSherry, F., Nissim, K. and Smith, A. (2006). Calibrating Noise to Sensitivity in Private Data Analysis, *Proceedings of the Third conference on Theory of Cryptography*, Springer-Verlag, pp. 265–284.

Dwork, C., McSherry, F., Nissim, K. and Smith, A. (2017). Calibrating noise to

sensitivity in private data analysis, *Journal of Privacy and Confidentiality* **7**(3): 17–51.

Dwork, C. and Roth, A. (2014). The Algorithmic Foundations of Differential Privacy, *Foundations and Trends in Theoretical Computer Science* **9**(3-4): 211–407.

Erlingsson, Ú., Pihur, V. and Korolova, A. (2014). RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response, *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security - CCS '14* pp. 1054–1067.

Fellegi, I. P. (1972). On the question of statistical confidentiality, *Journal of the American Statistical Association* **67**(337): 7–18.

Ganta, S. R., Kasiviswanathan, S. P. and Smith, A. (2008). Composition attacks and auxiliary information in data privacy, *KDD*.

Ghosh, A. and Roth, A. (2015). Selling privacy at auction, *Games and Economic Behavior* **91**: 334–346.

Goldfarb, A., Greenstein, S. M. and Tucker, C. E. (2015). *Economic Analysis of the Digital Economy*, University of Chicago Press.

Google (2019). Tensorflow privacy. https://github.com/tensorflow/privacy.

Gupta, A., Roth, A. and Ullman, J. (2012). Iterative constructions and private data release, *Proceedings of the 9th International Conference on Theory of Cryptography*, TCC'12, Springer-Verlag, Berlin, Heidelberg, pp. 339–356.

Hardt, M. and Rothblum, G. N. (2010). A multiplicative weights mechanism for privacy-preserving data analysis, *2010 IEEE 51st Annual Symposium on Foundations of Computer Science* pp. 61–70.

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

Kasiviswanathan, S. P. and Smith, A. (2014). On the 'semantics' of differential privacy: A bayesian formulation, *Journal of Privacy and Confidentiality* **6**(1): 1.

Kifer, D. and Lin, B.-R. (2010). Towards an axiomatization of statistical privacy and utility, *Proceedings of the Twenty-ninth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, PODS '10.

Kifer, D. and Machanavajjhala, A. (2011). No free lunch in data privacy, *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data*, SIGMOD '11, ACM Digital Library, New York, NY, USA, pp. 193–204.

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

Kifer, D. and Machanavajjhala, A. (2014). Pufferfish: A framework for mathematical privacy definitions, *ACM Trans. Database Syst.* **39**(1): 3.

Li, C., Hay, M., Rastogi, V., Miklau, G. and McGregor, A. (2010). Optimizing linear counting queries under differential privacy, *Proceedings of the twenty-ninth ACM*

SIGMOD-SIGACT-SIGART symposium on Principles of database systems of data - PODS10, Association for Computing Machinery (ACM), pp. 123–134.

Li, C., Li, D. Y., Miklau, G. and Suciu, D. A. N. (2014). A Theory of Pricing Private Data, *ACM Transactions on Database Systems* **39**(4): 34:1–34:27. Pages 34:1–34:27.

Machanavajjhala, A., Kifer, D., Abowd, J., Gehrke, J. and Vilhuber, L. (2008). Privacy: theory meets practice on the map, *Proceedings of the 2008 IEEE 24th International Conference on Data Engineering* pp. 277–286.

Mas-Colell, A., Whinston, M. and Green, J. (1995). *Microeconomic theory*, Oxford student edition, Oxford University Press.

McKenna, R., Miklau, G., Hay, M. and Machanavajjhala, A. (2018). Optimizing error of high-dimensional statistical queries under differential privacy, *Proceedings of the VLDB Endowment* **11**(10).

McMahan, H. B. and Andrew, G. (2018). A general approach to adding differential privacy to iterative training procedures, *CoRR* **abs/1812.06210**.

McSherry, F. D. (2009). Privacy integrated queries: An extensible platform for privacy-preserving data analysis, *Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data*.

Nissim, K., Orlandi, C. and Smorodinsky, R. (2012). Privacy-aware mechanism design, *EC '12 Proceedings of the 13th ACM Conference on Electronic Commerce* pp. 774–789.

Panel on Statistics on Natural Gas (1985). Natural gas needs in a changing regulatory environment, *National Academy Press* .

Ruggles, S., Fitch, C., Magnuson, D. and Schroeder, J. (2019). Differential privacy and census data: Implications for social and economic research, *AEA Papers and Proceedings* **109**: 403–08.

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

Spence, A. M. (1975). Monopoly, quality, and regulation, *The Bell Journal of Economics* **6**(2): 417–429.

Spencer, B. D. (1985). Optimal data quality, *Journal of the American Statistical Association* **80**(391): 564–573.

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).

US Census Bureau (2018). Post-Secondary Employment Outcomes (PSEO) (Beta). https://lehd.ces.census.gov/data/.

Warner, S. L. (1965). Randomised response: a survey technique for eliminating evasive answer bias, *Journal of the American Statistical Association* **60**(309): 63––69.

Wasserman, L. and Zhou, S. (2010). A Statistical Framework for Differential Privacy, *Journal of the American Statistical Association* **105**(489): 375–389.

Willenborg, L. and de Waal, T. (1996). *Statistical Disclosure Control in Practice*, Springer-Verlag New York.

Wood, A., Altman, M., Bembenek, A., Bun, M., Gaboardi, M., Honaker, J., Nissim, K., O'Brien, D. R., Steinke, T. and Vadhan, S. (2018). Differential privacy: A primer for a non-technical audience, *Vanderbilt Journal of Entertainment and Technology Law* **21**(1).