Why the Economics Profession Must Actively Participate in the Privacy Protection Debate

By JOHN M. ABOWD, IAN M. SCHMUTTE, WILLIAM N. SEXTON, AND LARS VILHUBER

Draft: January 2019

This version: January 19, 2019

Abstract: *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. 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.*

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 Commit-

* Abowd, Sexton, and Vilhuber: U.S. Census Bureau, Washington, DC, USA and Cornell University, Ives Hall, Ithaca, NY, USA. Schmutte: U.S. Census Bureau and University of Georgia, Amos Hall, Athens, GA, USA. This research was partially funded through NSF Grant #1131848 (NCRN) and Alfred P. Sloan Foundation Grant G-2015-13903. The views expressed in this paper are those of the authors and not those of the U.S. Census Bureau or other sponsors.

tee (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.

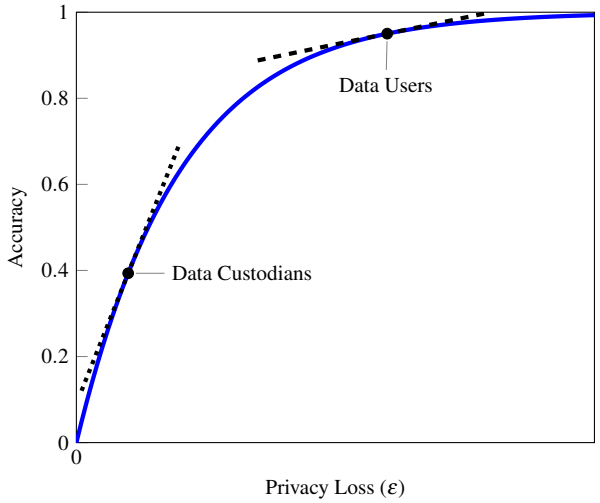

Figure 1. : The trade-off between privacy loss and accuracy in data publication

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 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). 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. 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, Pihur and Korolova, 2014), Microsoft (Ding, Kulkarni and Yekhanin, 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 science 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.

I. 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.

II. The Roles to be Played by Economists

Amid the sea change in the way confidential data are made available for research, economists have two 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.

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, Orlandi and Smorodinsky, 2012), as well as a growing evidence base for the private costs of privacy loss (e.g. Acquisti, John and Loewenstein, 2013).

III. 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, 2018). 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 microdata, 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 vague, 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

[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.

should aspire to do better. We should aspire to procedures that are provably inference valid.

IV. 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), which provides inference-valid tract-level summaries of inter-generational mobility by race and gender—an outcome that is not feasible using traditional SDL.

V. 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. 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.

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 $e^{2\varepsilon}$ (see the Online Appendix for details). 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, $\varepsilon = 1.0$ 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 $\varepsilon = 0.25$ guarantees that the improvement in the odds never exceeds $1.65:1$, and an $\varepsilon = 0.1$ 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).

VI. 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, Stanley and Totty, 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.

REFERENCES

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

Abowd, John M., and Ian M. Schmutte. 2019. "An Economic Analysis of Privacy Protection and Statistical Accuracy as Social Choices." *American Economic Review*, 109(1): 171–202.

Abowd, John M., Bryce E. Stephens, Lars Vilhuber, Fredrik Andersson, Kevin L. McKinney, Marc Roemer, and Simon D. Woodcock. 2009. "The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators." *Producer Dynamics: New Evidence from Micro Data*, , ed. Timothy Dunne, J. Bradford Jensen and Mark J. Roberts. University of Chicago Press.

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

Benedetto, Gary, Jordan C. Stanley, and Evan Totty. 2018. "The Creation and Use of the SIPP Synthetic Beta v7.0." U.S. Census Bureau.

Chetty, Raj, John Friedman, Nathaniel Hendren, Maggie Jones, and Sonya Porter. 2018. "The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility." National Bureau of Economic Research Working Paper w25147, National Bureau of Economic Research.

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

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

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

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

Dwork, Cynthia, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. "Calibrating Noise to Sensitivity in Private Data Analysis." *Tcc*, 265–284. DOI:10.1007/11681878_14.

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

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

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

Nissim, Kobbi, Claudio Orlandi, and Rann Smorodinsky. 2012. "Privacy-aware mechanism design." *EC '12 Proceedings of the 13th ACM Conference on Electronic Commerce*, 774–789.

Panel on Statistics on Natural Gas. 1985. "Natural Gas Needs in a Changing Regulatory Environment." *National Academy Press*.

Ruggles, Stephen. 2018. "Implications of Differential Privacy for Census Bureau Data and Scientific Research." Minnesota Population Center 2018-6, Version 5.1.

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

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

Wood, Alexandra, Micah Altman, Aaron Bembenek, Mark Bun, Marco Gaboardi, James Honaker, Kobbi Nissim, David R. O'Brien, Thomas Steinke, and Salil Vadhan. 2018. "Differential Privacy: A Primer for a Non-Technical Audience." *Vanderbilt Journal of Entertainment and Technology Law*, 21(1).

ONLINE APPENDIX

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 $\mathscr{D}$. 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,

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

Note that

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

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,

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

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:

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

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:

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

$$= \frac{\sum_{z\in\mathscr{D}}\mu(R=r,E,z)Pr[M(z)=\omega]/\sum_{z\in\mathscr{D}}\mu(R=r,E,z)Pr[M(\tilde{z})=\omega]}{\sum_{y\in\mathscr{D}}\mu(E,y)Pr[M(y)=\omega]/\sum_{y\in\mathscr{D}}\mu(E,y)Pr[M(\tilde{y})=\omega]}$$

$$\leq \frac{\sum_{z\in\mathscr{D}}\mu(R=r,E,z)e^{\varepsilon}Pr[M(\tilde{z})=\omega]/\sum_{z\in\mathscr{D}}\mu(R=r,E,z)Pr[M(\tilde{z})=\omega]}{\sum_{y\in\mathscr{D}}\mu(E,y)Pr[M(y)=\omega]/\sum_{y\in\mathscr{D}}\mu(E,y)Pr[M(\tilde{y})=\omega]}$$

(M is bounded ε-differentially private so $Pr[M(z)=\omega] \leq e^{\varepsilon}Pr[M(\tilde{z})=\omega]$.)

$$= \frac{e^{\varepsilon}\sum_{z\in\mathscr{D}}\mu(R=r,E,z)Pr[M(\tilde{z})=\omega]/\sum_{z\in\mathscr{D}}\mu(R=r,E,z)Pr[M(\tilde{z})=\omega]}{\sum_{y\in\mathscr{D}}\mu(E,y)Pr[M(y)=\omega]/\sum_{y\in\mathscr{D}}\mu(E,y)Pr[M(\tilde{y})=\omega]}$$

(Factor out e^{ε}.)

$$= \frac{e^{\varepsilon}}{\sum_{y\in\mathscr{D}}\mu(E,y)Pr[M(y)=\omega]/\sum_{y\in\mathscr{D}}\mu(E,y)Pr[M(\tilde{y})=\omega]}$$

(The summations in the numerator ratio cancel out; i.e., the ratio equals 1.)

$$\leq \frac{e^{\varepsilon}}{\sum_{y\in\mathscr{D}}\mu(E,y)e^{-\varepsilon}Pr[M(\tilde{y})=\omega]/\sum_{y\in\mathscr{D}}\mu(E,y)Pr[M(\tilde{y})=\omega]}$$

(M is bounded ε-differentially private so $Pr[M(y)=\omega] \geq e^{-\varepsilon}Pr[M(\tilde{y})=\omega]$.)

$$= \frac{e^{\varepsilon}}{e^{-\varepsilon}\sum_{y\in\mathscr{D}}\mu(E,y)Pr[M(\tilde{y})=\omega]/\sum_{y\in\mathscr{D}}\mu(E,y)Pr[M(\tilde{y})=\omega]}$$

(Factor out $e^{-\varepsilon}$)

$$= e^{2\varepsilon}$$

(The summations in the denominator ratio cancel out; i.e., the ratio equals 1.)

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