Privacy Elasticity: A (Hopefully) Useful New Concept

Inbal Dekel Rachel Cummings Ori Heffetz Katrina Ligett*

August 13, 2024

Abstract

Privacy considerations and their effects on behavior are becoming increasingly important. Yet the extremes of full and no privacy are rarely an option. How much does behavior change with small changes in privacy? Dekel et al. (2023) introduce the concept of *privacy elasticity*, the responsiveness of economic variables to small changes in privacy protections. This concept combines elasticity—a key economic measure of responsiveness of one variable to changes in another—and *differential privacy*—a computer science theory emerging as the standard tool for protecting and quantifying privacy. Together, they create a measure of privacy elasticity that is portable and comparable across contexts. The applicability of this concept is demonstrated by reviewing how privacy elasticity can be estimated in a public-good lab experiment.

KEYWORDS: privacy elasticity, differential privacy, privacy guarantees, visibility, economic experiments, public-good game. JEL CLASSIFICATION: C91, D82, Z00

*Dekel: Bogen Family Department of Economics, The Hebrew University of Jerusalem (e-mail: inbal.dekel1@mail.huji.ac.il). Cummings: Department of Industrial Engineering and Operations Research, Department of Computer Science, and Data Science Institute, Columbia University (e-mail: rac2239@columbia.edu). Heffetz: S.C. Johnson Graduate School of Management, Cornell University, Bogen Family Department of Economics and Federmann Rationality Center, The Hebrew University of Jerusalem, and NBER (e-mail: oh33@cornell.edu). Ligett: School of Computer Science and Engineering and Federmann Rationality Center, The Hebrew University of Jerusalem (e-mail: katrina.ligett@mail.huji.ac.il). For financial support, we acknowledge the following. Dekel and Ligett: The United States Air Force and DARPA under contracts FA8750-16-C-0022 and FA8750-19-2-0222, and the Federmann Cyber Security Center in conjunction with the Israel national cyber directorate. Cummings: NSF grants CNS-1850187 and CNS-1942772 (CAREER), the Defense Advanced Research Projects Agency under contract number W911NF-21-1-0371, a Mozilla Research Grant, a JPMorgan Chase Faculty Research Award, a Google Research Fellowship, and an Apple Privacy-Preserving Machine Learning Award; part of this work was completed while Cummings was affiliated with the California Institute of Technology and the Georgia Institute of Technology, and while visiting the Simons Institute for the Theory of Computing and the Hebrew University of Jerusalem. Heffetz: Israel Science Foundation (grant no. 1680/16), and the S.C. Johnson Graduate School of Management; part of this work was completed while Heffetz was visiting the School for Public and International Affairs (SPIA) at Princeton University. Ligett: NSF grants CNS-1254169 and CNS-1518941, US-Israel Binational Science Foundation grant 2012348, Israel Science Foundation (ISF) grant #1044/16, Israeli Science Foundation (ISF) grant #2861/20, a Google Faculty Research award, Simons Foundation Mathematical and Physical Science Collaboration number 733792, a gift to the McCourt School of Public Policy and Georgetown University, and a grant from the Israeli National Data Science initiative; part of this work was done while Ligett was visiting the Simons Institute for the Theory of Computing.

1 Introduction

Privacy considerations and their effects on behavior are becoming increasingly important. Yet, individuals rarely experience either full privacy or a complete lack of privacy. How much does behavior change with small changes in privacy? Our paper, "The Privacy Elasticity of Behavior: Conceptualization and Application" (Dekel et al., 2023) introduces the concept of privacy elasticity: the responsiveness of economic variables to small changes in privacy protections. This concept combines elasticity—a key economic measure of responsiveness, calculated by the percentage change in one variable resulting from a one-percent change in another—and *ϵ-differential privacy* (Dwork et al., 2006)—a computer science theory widely adopted by industry and government as a standard tool for protecting and quantifying continuous changes in privacy. Combining the two allows for a privacy elasticity measure that is portable and comparable across contexts.

In Section 2 we review the definition of differential privacy and our implied definition of privacy elasticity. Intuitively, differential privacy protects the privacy of individual data by adding random noise that limits the inferences that observers can make about an individual. While leading social scientists have reservations about this measure (Hotz et al., 2022), it is widely adopted by statistical agencies and tech companies, making it a natural starting point for measuring privacy elasticity. To demonstrate the usefulness of this concept, we review real-world deployments of differential privacy (Section 3) and a stylized theoretical example (Section 4) from our paper. Additionally, Section 5 reviews how privacy elasticity can be empirically estimated using a controlled public-good lab experiment, where noisy versions of subjects' contributions to the public good are announced to their group. Finally, Section 6 discusses some implications of our proposed notion of privacy elasticity.

2 Conceptualizing Privacy Elasticity

2.1 Differential Privacy

Differential privacy protects privacy by adding randomness to sensitive data, computations on such data, or the published results of such computations. This makes it difficult for observers of the released output to make inferences about any individual. Even if the output reveals something about an individual, similar inferences would be made if the individual's data were replaced with any other possible data, thus preserving their *differential* privacy. For more details, see Dwork and Roth (2014) or Heffetz and Ligett (2014).

Consider some sensitive personal data that an individual may not want disclosed to others (e.g., researchers, Silicon Valley companies, the government, or the public). Imagine that some computation is performed on this data, whose result will be published or recorded in some database. An ϵ-locally differentially private (Dwork et al., 2006) computation selects its output (i.e., the published or recorded signal) using a degree of randomness, such that any given output is *similarly likely* under any two possible data profiles of an individual.

Specifically, the likelihood of any output under any two possible data profiles differs by at most a fixed multiplier, e^ϵ, where ϵ is a non-negative parameter quantifying the level of privacy. The smaller ϵ is, the less the output depends on the specific data profile used, offering stronger privacy but less accurate results. When $\epsilon = 0$ then $e^\epsilon = 1$ and the output distribution is the same regardless of the data profile used, providing perfect privacy, but a perfectly useless signal. When $\epsilon = \infty$ there is no privacy, but the signal is perfectly informative. In between, there is a continuum where the tradeoff between privacy and accuracy is transparent and measurable.

2.2 Privacy Elasticity

Elasticity is a key concept in economics. It measures the percentage change in an economic variable of interest in response to a one-percent change in another variable. In essence, it measures the sensitivity of one variable to (small) changes in another. It is often used to measure how quantities demanded and supplied respond to changes in price, income, and other factors. For instance, a government that considers raising the tax on a good needs to take into account how much the demand will fall due to the price increase, which requires understanding the price elasticity of demand for that good. Importantly, elasticity does not rely on the units of the variables involved. However, it does require a measure for the variables involved where small, proportional changes are meaningful.

To define privacy elasticity—the percentage change in a variable in response to a one-percent change in privacy—we need a privacy metric that meets these criteria. Our paper suggests using e^ϵ, the maximum proportional increase in the likelihood of disclosing information about an individual when using their own data compared to any other possible data. Hotz et al. (2022) argue against widespread use of differential privacy, as it measures the risk of disclosing information relatively (proportionately) rather than additively (absolutely). They argue that individuals may be more concerned about small increases in risk if the initial risk is high, and less concerned about larger increases if the initial risk is low. These concerns notwithstanding, differential privacy's multiplicative (relative, proportional) approach is already in widespread use. Our paper could thus be viewed as conceptualizing and measuring the extent to which people do respond to this imperfect but influential lever. Moreover, using our approach to estimate privacy elasticity when the baseline risk is known could contribute to the discussion by providing empirical evidence relevant to their argument.

To fix ideas, imagine a person participating in an activity on a platform, such as responding to a government survey, browsing the web, or contributing to a public good. This person might be concerned that if certain actions (e.g., survey responses, websites visited, or public-good contributions) were tracked or revealed, it might increase the likelihood of bad outcomes, such as online targeting or social repercussions. If the platform uses ϵ-differential privacy then the person is guaranteed that no matter which action she takes, the information reported about her will be almost the same. This means the likelihood of bad outcomes will also not change much. Specifically, the probability of any reportable information and its potential implications won't increase by more than a constant factor, e^ϵ, compared to the lowest possible probability.

A one-percent increase in privacy loss in this setting means a one-percent increase in e^ϵ used by the platform. This implies the concept of privacy elasticity, defined as the percentage change in a variable in response to a one-percent change in the upper bound on the ratio between the probability of any outcome induced by the privacy mechanism and what it would have been if an individual's actions were entirely different. That is, privacy elasticity measures how much a behavior changes when there is a small change in privacy protection.

3 Potential Applications

We now review real-world deployments of differential privacy cited in our paper, where privacy elasticity may be useful.

For instance, Apple Watch users can use an ECG app to record heartbeats and check for irregular heart rhythms. Individuals might worry that merely using the ECG app might imply a heart condition, which if revealed to others might increase the probability of some adverse treatment by insurers, advertisers, employers or even potential romantic partners. Apple currently uses $\epsilon = 2$ per day to protect this information before it is gathered. If changes in privacy guarantees affect users' willingness to use the ECG feature or share health data with Apple, this could have important implications, from affecting Apple's ability to do strategic product planning, to making the Apple Watch a less useful product, to even saving lives. Understanding the privacy elasticity of such behavioral changes could be important.

As another example, both Google and Apple use differential privacy to protect and gather information on user web-browsing behavior. Users might be concerned that visiting certain websites might reveal sensitive or embarrassing information about them. Hence, they might alter their browsing behavior or their willingness to share browsing data with tech companies based on the level of privacy guaranteed. Therefore, understanding the privacy elasticity of behavior in these settings could be useful. For instance, it could help companies balance users' privacy with their need to fix browser crashes.

There are many other examples, some using more advanced differential privacy techniques than we cover in our paper. Windows has used differential privacy to protect data collected from millions of Windows 10 devices about users' app usage and to protect information that it reveals to managers about employee collaborations. Microsoft also uses differential privacy for advertiser queries on LinkedIn, and for Microsoft Office's suggested email replies. Other major tech companies, including Uber, Snapchat, Salesforce, Facebook, and Amazon, are developing or implementing tools for differentially private data analysis, and TikTok has posted job ads that describe background in differential privacy as a qualification.

4 A Stylized Example

To make these potential real-world applications of privacy elasticity more concrete, consider the following scenario analyzed in our paper. A mobile-device firm wishes to estimate the fraction of its shipped devices that are defective. Users cannot detect the defect themselves, so the firm asks each owner to run a simple diagnostic test. This test perfectly detects device status, and sends a (possibly noisy) signal to the firm. Users decide whether to opt in or out of running the diagnostic. Before running it, each device is equally likely to be defective, so a user's opt-in decision is independent of their device status.

Individuals may not want their device status to be revealed to the firm, as they might worry about unknown future implications and, more generally, tech-firms' use of their private data. They are therefore more likely to opt-in if they are guaranteed more privacy. Hence, the firm embeds the diagnostic within an ϵ-locally differentialy private mechanism. The firm's optimization problem is to choose the level of privacy (e^ϵ) to maximize the accuracy (i.e., reduce the variance) of the estimator of the fraction.

The firm faces a privacy-accuracy tradeoff: the more privacy it offers (i.e., the lower it sets e^ϵ), the more individuals will be willing to opt in, increasing the sample size and thus reducing the variance of the estimator. On the other hand, the more privacy there is, the noisier (less accurate) any sent signal becomes, which increases the estimator's variance. Which effect is stronger depends on the privacy elasticity of total participation.

Our paper shows that if the firm's only goal is to maximize data accuracy, it will only offer privacy if privacy elasticity is very high (such that the increase in participation outweighs the added noise due to the increase in privacy). This highlights the importance of privacy elasticity in economic decisions. In many real-world cases, privacy elasticity may currently be low. Starting with a socially desirable privacy level, privacy elasticity can be used to measure the change in user behavior needed to incentivize firms to provide at least that level of privacy. If achieving sufficiently high elasticity seems unlikely, then regulatory intervention may increase the likelihood of securing the desired privacy level. Moreover, with time and experience, individuals may come to correctly interpret noisy signals, and this may increase the privacy elasticity of participation.

5 Privacy Elasticity in a Public-Good Experiment

To demonstrate how to measure privacy elasticity, our paper incorporates a differentially private announcement mechanism into a public-good-game lab experiment.

In the experiment, 328 subjects play seven rounds of a public-good game in fixed groups of eight, seated in the same computer lab. In each round, each subject divides a \$10 endowment between a personal account and a group account. Allocating money to the personal account yields a higher individual return, but the group benefits more from contributions to the group account, which are first multiplied by some multiplier, and then redistributed among the subject and others. We randomly vary across sessions the price of generating \$1 in others' money.

To incentivize subjects to play in each round as if it were the only round that mattered, they are informed in advance that one round will be randomly chosen at the end of the experiment, and their payments will be based on the allocations from that round. Additionally, noisy versions of these allocations will be announced to their group. The level of noise is randomly varied across session rounds. Announcements are made by having each subject's noisy allocation in the chosen round both appear on everyone's screen and read aloud by an experimenter while the subject stands up and faces the group. Subjects are not told in advance about the varying privacy levels. The experiment is double-blind, with payments prepared in a different room by another experimenter.

We estimate an average price elasticity of contribution at -0.23 (S.E. = 0.07). Furthermore, we estimate an average privacy elasticity of contribution (over our finite ϵ's) at 0.07 (S.E. = 0.01). Importantly, this privacy elasticity is not a mere reaction of subjects to *changes* in privacy levels. The between-subjects estimate (based on the first round of each session, where subjects did not know that they would face other privacy levels) is close, at 0.06 (S.E. = 0.03). These elasticities suggest that contributions are similarly affected by a one-percent increase in price and a 3–4 percent increase in privacy. They also suggest that contributions are *inelastic* with respect to both price and, even more so, privacy: when either price or privacy changes by one percent, contributions change by *less* than one percent, specifically by only 0.23 or 0.06–0.07 percent, respectively.

6 Discussion

Personal data is collected and stored at an ever increasing pace, raising the importance of privacy concerns and their potential effects on behavior. Since privacy is rarely binary, privacy elasticity may help understand how privacy protections affect economic systems. Differential privacy, the emerging standard for privacy protections in large data systems, offers a tool for quantifying small changes in privacy. While we hope future work will explore other notions of privacy, at present we define privacy elasticity using this imperfect, yet widely used, measure.

Currently, many individuals are likely unaware of the level of privacy protecting their data and of its potential consequences. As a result, behavior may be inelastic to privacy in many real-world settings. However, as privacy awareness grows through education, legislation, and regulation, privacy elasticity may significantly increase.

Abowd and Schmutte (2019) discuss the tradeoff faced by statistical agencies when deciding how much noise to add to published statistics. More noise enhances respondent privacy but reduces the accuracy of the statistics. They call for research to find the optimal balance. Our work shows that finding this balance is more complex than simply choosing ϵ along a fixed tradeoff curve. Different privacy levels can change behavior and participation in a dataset, affecting the data itself. Thus, unless behavior is completely privacy inelastic, the choice of ϵ can have complex effects on the data gathered, its accuracy, and its representativeness.

References

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.

Dekel, Inbal, Rachel Cummings, Ori Heffetz, and Katrina Ligett. 2023. “The privacy elasticity of behavior: Conceptualization and application.” NBER Working Paper w30215.

Dwork, Cynthia, and Aaron Roth. 2014. “The algorithmic foundations of differential privacy.” *Foundations and Trends in Theoretical Computer Science*, 9(3–4): 211–407.

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

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

Hotz, V. Joseph, Christopher R. Bollinger, Tatiana Komarova, Charles F. Manski, Robert A. Moffitt, Denis Nekipelov, Aaron Sojourner, and Bruce D. Spencer. 2022. “Balancing data privacy and usability in the federal statistical system.” *Proceedings of the National Academy of Sciences*, 119(31): e2104906119.