Differential Privacy: A Primer for a Non-Technical Audience

Differential Privacy: A Primer for a Non-Technical Audience

*Alexandra Wood, Micah Altman, Aaron Bembenek, Mark Bun, Marco Gaboardi, James Honaker, Kobbi Nissim, David R. O'Brien, Thomas Steinke & Salil Vadhan**

ABSTRACT

Differential privacy is a formal mathematical framework for quantifying and managing privacy risks. It provides provable privacy protection against a wide range of potential attacks, including those

* Alexandra Wood is a Fellow at the Berkman Klein Center for Internet & Society at Harvard University. Micah Altman is Director of Research at MIT Libraries. Aaron Bembenek is a PhD student in computer science at Harvard University. Mark Bun is a Google Research Fellow at the Simons Institute for the Theory of Computing. Marco Gaboardi is an Assistant Professor in the Computer Science and Engineering department at the State University of New York at Buffalo. James Honaker is a Research Associate at the Center for Research on Computation and Society at the Harvard John A. Paulson School of Engineering and Applied Sciences. Kobbi Nissim is a McDevitt Chair in Computer Science at Georgetown University and an Affiliate Professor at Georgetown University Law Center; work towards this document was completed in part while the Author was visiting the Center for Research on Computation and Society at Harvard University. David R. O'Brien is a Senior Researcher at the Berkman Klein Center for Internet & Society at Harvard University. Thomas Steinke is a Research Staff Member at IBM Research – Almaden. Salil Vadhan is the Vicky Joseph Professor of Computer Science and Applied Mathematics at Harvard University.

This Article is the product of a working group of the *Privacy Tools for Sharing Research Data* project at Harvard University (http://privacytools.seas.harvard.edu). The working group discussions were led by Kobbi Nissim. Alexandra Wood and Kobbi Nissim are the lead Authors of this Article. Working group members Micah Altman, Aaron Bembenek, Mark Bun, Marco Gaboardi, James Honaker, Kobbi Nissim, David R. O'Brien, Thomas Steinke, Salil Vadhan, and Alexandra Wood contributed to the conception of the Article and to the writing. The Authors thank John Abowd, Scott Bradner, Cynthia Dwork, Simson Garfinkel, Caper Gooden, Deborah Hurley, Rachel Kalmar, Georgios Kellaris, Daniel Muise, Michel Reymond, and Michael Washington for their many valuable comments on earlier versions of this Article. A preliminary version of this work was presented at the 9th Annual Privacy Law Scholars Conference (PLSC 2017), and the Authors thank the participants for contributing thoughtful feedback. The original manuscript was based upon work supported by the National Science Foundation under Grant No. CNS-1237235, as well as by the Alfred P. Sloan Foundation. The Authors' subsequent revisions to the manuscript were supported, in part, by the US Census Bureau under cooperative agreement no. CB16ADR0160001. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the Authors and do not necessarily reflect the views of the National Science Foundation, the Alfred P. Sloan Foundation, or the US Census Bureau.

currently unforeseen. Differential privacy is primarily studied in the context of the collection, analysis, and release of aggregate statistics. These range from simple statistical estimations, such as averages, to machine learning. Tools for differentially private analysis are now in early stages of implementation and use across a variety of academic, industry, and government settings. Interest in the concept is growing among potential users of the tools, as well as within legal and policy communities, as it holds promise as a potential approach to satisfying legal requirements for privacy protection when handling personal information. In particular, differential privacy may be seen as a technical solution for analyzing and sharing data while protecting the privacy of individuals in accordance with existing legal or policy requirements for de-identification or disclosure limitation.

This primer seeks to introduce the concept of differential privacy and its privacy implications to non-technical audiences. It provides a simplified and informal, but mathematically accurate, description of differential privacy. Using intuitive illustrations and limited mathematical formalism, it discusses the definition of differential privacy, how differential privacy addresses privacy risks, how differentially private analyses are constructed, and how such analyses can be used in practice. A series of illustrations is used to show how practitioners and policymakers can conceptualize the guarantees provided by differential privacy. These illustrations are also used to explain related concepts, such as composition (the accumulation of risk across multiple analyses), privacy loss parameters, and privacy budgets. This primer aims to provide a foundation that can guide future decisions when analyzing and sharing statistical data about individuals, informing individuals about the privacy protection they will be afforded, and designing policies and regulations for robust privacy protection.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Differential privacy is a strong, mathematical definition of privacy in the context of statistical and machine learning analysis. It is used to enable the collection, analysis, and sharing of a broad range of statistical estimates based on personal data, such as averages, contingency tables, and synthetic data, while protecting the privacy of the individuals in the data.

Differential privacy is not a single tool, but rather a criterion, which many tools for analyzing sensitive personal information have been devised to satisfy. It provides a mathematically provable guarantee of privacy protection against a wide range of *privacy attacks*, defined as attempts to learn private information specific to individuals from a data release. Privacy attacks include re-identification, record linkage, and differencing attacks, but may also include other attacks currently unknown or unforeseen. These concerns are separate from security attacks, which are characterized by attempts to exploit vulnerabilities in order to gain unauthorized access to a system.

Computer scientists have developed a robust theory for differential privacy over the last fifteen years, and major commercial and government implementations are starting to emerge.

The differential privacy guarantee (Part III). Differential privacy mathematically guarantees that anyone viewing the result of a differentially private analysis will essentially make the same inference about any individual's private information, whether or not that individual's private information is included in the input to the analysis.

The privacy loss parameter (Section IV.B). What can be learned about an individual as a result of her private information being included in a differentially private analysis is limited and quantified by a privacy loss parameter, usually denoted epsilon (ε). Privacy loss can grow as an individual's information is used in multiple analyses, but the increase is bounded as a known function of ε and the number of analyses performed.

Interpreting the guarantee (Section VI.C). The differential privacy guarantee can be understood in reference to other privacy concepts:

- Differential privacy protects an individual's information essentially as if her information were not used in the analysis at all, in the sense that the outcome of a differentially private algorithm is approximately the same whether the individual's information was used or not.
- Differential privacy ensures that using an individual's data will not reveal essentially any personally identifiable information that is specific to her, or even whether the individual's information was used at all. Here, *specific* refers to information that cannot be inferred unless the individual's information is used in the analysis.

As these statements suggest, differential privacy is a new way of protecting privacy that is more quantifiable and comprehensive than the concepts of privacy underlying many existing laws, policies, and practices around privacy and data protection. The differential privacy

guarantee can be interpreted in reference to these other concepts, and can even accommodate variations in how they are defined across different laws. In many settings, data holders may be able to use differential privacy to demonstrate that they have complied with applicable legal and policy requirements for privacy protection.

Differentially private tools (Part VII). Differential privacy is currently in initial stages of implementation and use in various academic, industry, and government settings, and the number of practical tools providing this guarantee is continually growing. Multiple implementations of differential privacy have been deployed by corporations such as Google, Apple, and Uber, as well as federal agencies like the US Census Bureau. Additional differentially private tools are currently under development across industry and academia.

Some differentially private tools utilize an interactive mechanism, enabling users to submit queries about a dataset and receive corresponding differentially private results, such as custom-generated linear regressions. Other tools are non-interactive, enabling static data or data summaries, such as synthetic data or contingency tables, to be released and used.

In addition, some tools rely on a curator model, in which a database administrator has access to and uses private data to generate differentially private data summaries. Others rely on a local model, which does not require individuals to share their private data with a trusted third party, but rather requires individuals to answer questions about their own data in a differentially private manner. In a local model, each of these differentially private answers is not useful on its own, but many of them can be aggregated to perform useful statistical analysis.

Benefits of differential privacy (Part VIII). Differential privacy is supported by a rich and rapidly advancing theory that enables one to reason with mathematical rigor about privacy risk. Adopting this formal approach to privacy yields a number of practical benefits for users:

- Systems that adhere to strong formal definitions like differential privacy provide protection that is robust to a wide range of potential privacy attacks, including attacks that are unknown at the time of deployment. An analyst using differentially private tools need not anticipate particular types of privacy attacks, as the guarantees of differential privacy hold regardless of the attack method that may be used.
- Differential privacy provides provable privacy guarantees with respect to the cumulative risk from successive data

releases and is the only existing approach to privacy that provides such a guarantee.

- Differentially private tools also have the benefit of transparency, as it is not necessary to maintain secrecy around a differentially private computation or its parameters. This feature distinguishes differentially private tools from traditional de-identification techniques, which often conceal the extent to which the data have been transformed, thereby leaving data users with uncertainty regarding the accuracy of analyses on the data.
- Differentially private tools can be used to provide broad, public access to data or data summaries while preserving privacy. They can even enable wide access to data that cannot otherwise be shared due to privacy concerns. An important example is the use of differentially private synthetic data generation to produce public-use microdata.

Differentially private tools can, therefore, help enable researchers, policymakers, and businesses to analyze and share sensitive data, while providing strong guarantees of privacy to the individuals in the data.

Keywords: differential privacy, data privacy, social science research

I. INTRODUCTION

Businesses, government agencies, and research institutions often use and share data containing sensitive or confidential information about individuals.[1] Improper disclosure of such data can have adverse consequences for a data subject's reputation, finances, employability, and insurability, as well as lead to civil liability, criminal penalties, or physical or emotional injuries.[2] Due to these issues and other related concerns, a large body of laws, regulations, ethical codes, institutional policies, contracts, and best practices has emerged to address potential privacy-related harms associated with the collection, use, and release of personal information.[3] The following discussion

1. *See* PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., EXEC. OFFICE OF THE PRESIDENT, BIG DATA AND PRIVACY: A TECHNOLOGICAL PERSPECTIVE (2014), https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf [https://perma.cc/MM2V-8C2P] (analyzing the current state of big-data collection, storage, and use in order to make policy recommendations).

2. *See generally* Daniel J. Solove, *A Taxonomy of Privacy*, 154 U. PA. L. REV. 477 (2006) (grouping different types of privacy violations and noting their potential harms).

3. *See* DANIEL J. SOLOVE & PAUL M. SCHWARTZ, INFORMATION PRIVACY LAW 2 (6th ed. 2018).

provides an overview of the broader data privacy landscape that has motivated the development of formal privacy models like differential privacy.

A. Introduction to Legal and Ethical Frameworks for Data Privacy

The legal framework for privacy protection in the United States has evolved as a patchwork of highly sector- and context-specific federal and state laws.[4] For instance, Congress has enacted federal information privacy laws to protect certain categories of personal information found in health,[5] education,[6] financial,[7] and government records,[8] among others. These laws often expressly protect information classified as personally identifiable information (PII), which generally refers to information that can be linked to an individual's identity or attributes.[9] Some laws also incorporate de-identification provisions, which provide for the release of information that has been stripped of PII.[10] State data protection and breach notification laws prescribe specific data security and breach reporting requirements when managing certain types of personal information.[11]

In addition, federal regulations generally require researchers conducting studies involving human subjects to secure approval from an institutional review board and fulfill ethical obligations to the participants, such as disclosing the risks of participation, obtaining their informed consent, and implementing specific measures to protect

4. *See id.* at 36–38.

5. *See, e.g.*, Health Insurance Portability and Accountability Act (HIPAA), Pub. L. No. 104-191, 110 Stat. 1936 (1996) (codified as amended in scattered titles of the U.S.C.).

6. *See, e.g.*, Family Educational Rights and Privacy Act of 1974 (FERPA), Pub. L. No. 93-380, 88 Stat. 571 (1974) (codified as amended at 20 U.S.C. § 1232g (2012)).

7. *See, e.g.*, Fair Credit Reporting Act, Pub. L. No. 91-508, 84 Stat. 1114 (1970) (codified at 15 U.S.C. §§ 1681–1681x); Gramm-Leach-Bliley Act, Pub. L. No. 106-102, 113 Stat. 1338 (1999) (codified in relevant part primarily at 15 U.S.C. §§ 6801–6809, §§ 6821–6827).

8. *See, e.g.*, Privacy Act of 1974, Pub. L. No. 93-579, 88 Stat. 1897 (1974) (codified as amended at 5 U.S.C. § 552a (2012)).

9. *See* SIMSON L. GARFINKEL, NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, DE-IDENTIFYING GOVERNMENT DATASETS 46, NIST Special Publication No. 800-188 (2d Draft, 2016), https://csrc.nist.gov/csrc/media/publications/sp/800-188/draft/documents/sp800_188_draft2.pdf [https://perma.cc/U6ZG-BFV5]; Paul M. Schwartz & Daniel J. Solove, *The PII Problem: Privacy and a New Concept of Personally Identifiable Information*, 86 N.Y.U. L. REV. 1814, 1816 (2011).

10. *See, e.g.*, DEP'T OF HEALTH & HUMAN SERVS., GUIDANCE REGARDING METHODS FOR DE-IDENTIFICATION OF PROTECTED HEALTH INFORMATION IN ACCORDANCE WITH THE HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT (HIPPA) PRIVACY RULE 6–7 (2012), https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/understanding/coveredentities/De-identification/hhs_deid_guidance.pdf [https://perma.cc/NRY2-M7J7].

11. *See* Paul M. Schwartz & Edward J. Janger, *Notification of Data Security Breaches*, 105 MICH. L. REV. 913, 972–74 (2007) (summarizing state security breach notification laws).

privacy.[12] It is also common for universities and other research institutions to adopt policies that require their faculty, staff, and students to abide by certain ethical and professional responsibility standards and set forth enforcement procedures and penalties for mishandling data.[13]

Further restrictions apply when privacy-sensitive data are shared under the terms of a data sharing agreement, which will often strictly limit how the recipient can use or redisclose the data received.[14] Organizations may also require privacy measures set forth by technical standards, such as those specifying information security controls to protect personally identifiable information.[15]

In addition, laws such as the EU General Data Protection Regulation are in place to protect personal data about European citizens regardless of where the data reside.[16] International privacy guidelines, such as the privacy principles developed by the Organisation for Economic Co-operation and Development, have also been adopted by governments across the world.[17] Moreover, the right to privacy is also protected by various international treaties and national constitutions.[18]

Taken together, the safeguards required by these legal and ethical frameworks are designed to protect the privacy of individuals and ensure they fully understand both the scope of personal information to be collected and the associated privacy risks. They also help data holders avoid administrative, civil, and criminal penalties, as well as maintain the public's trust and confidence in commercial, government, and research activities involving personal data.

---

12. *See* Protection of Human Subjects, 45 C.F.R. §§ 46.109, .111, .116 (2018).

13. *See, e.g.*, HARVARD UNIV. OFFICE OF THE VICE PROVOST FOR RESEARCH, HARVARD RESEARCH DATA SECURITY POLICY (2014), http://files.vpr.harvard.edu/files/vpr-documents/files/hrdsp_10_14_14_final_edits.pdf [https://perma.cc/BDW6-T5NF].

14. *See* ALEX KANOUS & ELAINE BROCK, INTER-UNIV. CONSORTIUM FOR POLITICAL & SOC. REFORM, CONTRACTUAL LIMITATIONS ON DATA SHARING 3 (2015), https://deepblue.lib.umich.edu/bitstream/handle/2027.42/123016/ContractualLimitationsonDataSharing150411-1.pdf [https://perma.cc/8JAQ-LWHP].

15. *See, e.g.*, INT'L ORG. FOR STANDARDIZATION, ISO 27018 CODE OF PRACTICE FOR PROTECTION OF PERSONALLY IDENTIFIABLE INFORMATION (PII) IN PUBLIC CLOUDS ACTING AS PII PROCESSORS (2014), https://www.iso.org/standard/61498.html [https://perma.cc/6R3L-SH3R] (abstract and preview).

16. *See* Commission Regulation 2016/679, 2016 O.J. (L 119) 1 [hereinafter GDPR].

17. *See* SOLOVE & SCHWARTZ, *supra* note 3, at 38. *See generally* Organisation for Economic Cooperation and Development [OECD], *Guidelines Governing the Protection of Privacy and Transborder Flow of Personal Data*, C(80)58 (July 11, 2013), https://www.oecd.org/sti/ieconomy/2013-oecd-privacy-guidelines.pdf [https://perma.cc/7SX3-ZEBP] (amending 1980 version).

18. *See, e.g.*, G.A. Res. 217 (III) A, Universal Declaration of Human Rights, art. 12 (Dec. 10, 1948).

B. Traditional Statistical Disclosure Limitation Techniques

A number of technical measures for disclosing data while protecting the privacy of individuals have been produced within the context of these legal and ethical frameworks.[19] In particular, statistical agencies, data analysts, and researchers have widely adopted a collection of statistical disclosure limitation (SDL) techniques to analyze and share data containing privacy-sensitive data with the aim of making it more difficult to learn personal information pertaining to an individual.[20] This category of techniques encompasses a wide range of methods for suppressing, aggregating, perturbing, and generalizing attributes of individuals in the data.[21] Such techniques are often applied with the explicit goal of de-identification—namely, making it difficult to link an identified person to a record in a data release by redacting or coarsening data.[22]

Advances in analytical capabilities, increases in computational power, and the expanding availability of personal data from a wide range of sources are eroding the effectiveness of traditional SDL techniques.[23] Since the 1990s—and with increasing frequency—privacy and security researchers have demonstrated that data that have been de-identified can often be successfully *re*-identified via a technique such as record linkage.[24] Re-identification via record linkage, or a linkage attack, refers to the re-identification of one or more records in a de-identified dataset by uniquely linking a record in a de-identified dataset with identified records in a publicly available dataset, such as a voter registration list.[25] As described in Example 1 below, in the late 1990s, Latanya Sweeney famously applied such an attack on a dataset containing de-identified hospital records.[26] Sweeney observed that records in the de-identified dataset contained the date of birth, sex, and

19. *See generally* Fed. Comm. on Statistical Methodology, *Report on Statistical Disclosure Limitation Methodology* (Office of Mgmt. & Budget: Statistical Policy, Working Paper No. 22, 2005), https://www.hhs.gov/sites/default/files/spwp22.pdf [https://perma.cc/LXN5-7QRQ].

20. *See id.* at 8.

21. *See id.* at 12–33 (describing various SDL techniques).

22. *See* GARFINKEL, *supra* note 9, at 3.

23. *See* Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1716, 1731, 1742 (2010).

24. *See id.* at 1719–22.

25. *See* CYNTHIA DWORK & AARON ROTH, THE ALGORITHMIC FOUNDATIONS OF DIFFERENTIAL PRIVACY 6–7 (2014) (originally published in 9 FOUND. & TRENDS IN THEORETICAL COMPUTER SCI. 211 (2014)); GARFINKEL, *supra* note 9, at 47.

26. *See Recommendations to Identify and Combat Privacy Problems in the Commonwealth: Hearing on H.R. 351 Before the H. Select Comm. on Information Security*, 189th Sess. (Pa. 2005) [hereinafter *Pa. Privacy Hearing*] (statement of Latanya Sweeney, Associate Professor, Carnegie Mellon University), http://dataprivacylab.org/dataprivacy/talks/Flick-05-10.html [https://perma.cc/W62P-Y2YX].

ZIP code of patients; that many of the patients had a unique combination of these three attributes; and that these three attributes were listed alongside individuals' names and addresses in publicly available voting records.[27] Sweeney used this information to re-identify records in the de-identified dataset.[28] Subsequent attacks on protected data have demonstrated weaknesses in other traditional approaches to privacy protection, and understanding the limits of these traditional techniques is the subject of ongoing research.[29]

C. The Emergence of Formal Privacy Models

Re-identification attacks are becoming increasingly sophisticated over time, as are other types of attacks that seek to infer characteristics of individuals based on information about them in a data set.[30] Successful attacks on de-identified data illustrate that traditional technical measures for privacy protection may be particularly vulnerable to attacks devised after a technique's deployment and use.[31] Some de-identification techniques, for example, require the specification of attributes in the data as identifying (e.g., names, dates of birth, or addresses) or non-identifying (e.g., movie ratings or hospital admission dates).[32] Data providers may later discover that attributes initially believed to be non-identifying can in fact be used to re-identify individuals.[33] Similarly, de-identification procedures may require a careful analysis of present and future data sources that could potentially be linked with the de-identified data and enable re-identification of the data. Anticipating the types of attacks and resources an attacker could leverage is a challenging exercise and ultimately will fail to address all potential attacks, as unanticipated

27. *See id.*

28. *See id.*

29. *See, e.g.*, Joseph A. Calandrino et al., *"You Might Also Like:" Privacy Risks of Collaborative Filtering*, 2011 IEEE SYMP. ON SECURITY & PRIVACY 231, 245; Yves-Alexandre de Montjoye et al., *Unique in the Crowd: The Privacy Bounds of Human Mobility*, NATURE SCI. REP. 4 (Mar. 25, 2013), https://www.nature.com/articles/srep01376.pdf [https://perma.cc/F8DZ-347V]; Arvind Narayanan & Vitaly Shmatikov, *Robust De-anonymization of Large Sparse Datasets*, 2008 IEEE SYMP. ON SECURITY & PRIVACY 111, 123–24.

30. *See, e.g.*, Irit Dinur & Kobbi Nissim, *Revealing Information While Preserving Privacy*, 22 PROC. ACM SIGMOD-SIGACT-SIGART SYMP. ON PRINCIPLES DATABASE SYS. 202, 203–04 (2003). *See generally* Arvind Narayanan, Joanna Huey & Edward W. Felten, *A Precautionary Approach to Big Data Privacy, in* DATA PROTECTION ON THE MOVE: CURRENT DEVELOPMENTS IN ICT AND PRIVACY/DATA PROTECTION 357 (Serge Gutwirth et al. eds., 2016).

31. *See* Narayanan, Huey & Felten, *supra* note 30, at 366.

32. *See* GARFINKEL, *supra* note 9, at 12, 38–40.

33. *See* Ohm, *supra* note 23, at 1723.

sources of auxiliary information that can be used for re-identification may become available in the future.[34]

Issues such as these underscore the need for privacy technologies that are immune not only to linkage attacks, but to any potential attack, including those currently unknown or unforeseen.[35] They also demonstrate that privacy technologies must provide meaningful privacy protection in settings where extensive external information may be available to potential attackers, such as employers, insurance companies, relatives, and friends of an individual in the data.[36] Real-world attacks further illustrate that ex post remedies, such as simply "taking the data back" when a vulnerability is discovered, are ineffective because many copies of a set of data typically exist, and copies often persist online indefinitely.[37]

In response to the accumulated evidence of weaknesses with respect to traditional approaches, a new privacy paradigm has emerged from the computer science literature—differential privacy.[38] Differential privacy is primarily studied in the context of the collection, analysis, and release of aggregate statistics. Such analyses range from simple statistical estimations—such as averages—to machine learning.[39] Contrary to common intuition, aggregate statistics such as these are not always safe to release because, as Part III explains, they can often be combined to reveal sensitive information about individual data subjects.

First presented in 2006,[40] differential privacy is the subject of ongoing research to develop privacy technologies that provide robust protection against a wide range of potential attacks.[41] Importantly, differential privacy is not a single tool but a definition or standard for

---

34. *See* GARFINKEL, *supra* note 9, at 38–40.

35. *See* Narayanan, Huey & Felten, *supra* note 30, at 370.

36. *See id.* at 362–63.

37. As an example, in 2006, AOL published anonymized search histories of over 650,000 users over a period of three months. Shortly after the release, journalists for the *New York Times* identified a person in the release, and AOL removed the data from its web site. *See* Michael Barbaro & Tom Zeller Jr., *A Face Is Exposed for AOL Searcher No. 4417749,* N.Y. TIMES (Aug. 9, 2006), https://www.nytimes.com/2006/08/09/technology/09aol.html [https://perma.cc/GWH2-W7F8]. However, in spite of AOL's withdrawal of the data, copies of the data are still accessible on the internet today. *See, e.g.*, *AOL Search Data Collection*, INTERNET ARCHIVE (Feb. 20, 2014), https://archive.org/details/AOL_search_data_leak_2006 [https://perma.cc/DVX3-KPUR].

38. *See generally* Cynthia Dwork et al., *Calibrating Noise to Sensitivity in Private Data Analysis,* 3 THEORY CRYPTOGRAPHY CONF. 265 (2006).

39. *See infra* Part V.

40. *See generally* Dwork et al., *supra* note 38.

41. *See, e.g.*, *Differential Privacy*, HARV. U. PRIVACY TOOLS PROJECT, https://privacytools.seas.harvard.edu/differential-privacy [https://perma.cc/FA7V-NZ3K] (last visited Sept. 14, 2018); *Putting Differential Privacy to Work*, U. PA., http://privacy.cis.upenn.edu [https://perma.cc/P5QU-XA7L] (last visited Sept. 14, 2018).

quantifying and managing privacy risks for which many technological tools have been devised.[42] Analyses performed with differential privacy differ from standard statistical analyses—such as the calculation of averages, medians, and linear regression equations—in that random noise[43] is added in the computation.[44] Tools for differentially private analysis are now in early stages of implementation and use across a variety of academic, industry, and government settings.[45]

This Article provides a simplified and informal, yet mathematically accurate, description of differential privacy.[46] Using intuitive illustrations and limited mathematical formalism, it describes the definition of differential privacy, how it addresses privacy risks, how differentially private analyses are constructed, and how such analyses can be used in practice. This discussion intends to help non-technical audiences understand the guarantees provided by differential privacy. It can help guide practitioners as they make decisions regarding whether to use differential privacy and, if so, what types of promises they should make to data subjects about the guarantees differential privacy provides. In addition, these illustrations intend to help legal scholars and policymakers consider how current and future legal frameworks and instruments should apply to tools based on formal privacy models such as differential privacy.

---

42. *See* Dwork et al., *supra* note 38, at 265; *infra* Part VII.

43. Random noise refers to uncertainty introduced into a computation by the addition of values sampled from a random process. For example, consider a computation that first calculates the number of individuals x in the dataset who suffer from diabetes, then samples a value y from a normal distribution with a mean of 0 and variance of 1, and outputs $z = x + y$. In this example, the random noise y is added in the computation to the exact count x to produce the noisy output z. For a more detailed explanation of random noise, see *infra* Part IV.

44. *See* Dwork et al., *supra* note 38, at 266.

45. *See infra* Part VII.

46. Differential privacy was defined in 2006 by Dwork, McSherry, Nissim and Smith. Dwork et al., *supra* note 38 (building on Avrim Blum et al., *Practical Privacy: The SuLQ Framework*, 24 PROC. ACM SIGMOD-SIGACT-SIGART SYMP. ON PRINCIPLES DATABASE SYS. 128, 128–30 (2005); Dinur & Nissim, *supra* note 30; Cynthia Dwork & Kobbi Nissim, *Privacy-Preserving Datamining on Vertically Partitioned Databases*, 24 ANN. INT'L CRYPTOLOGY CONF. 528 (2004); Alexandre Evfimievski, Johannes Gehrke, Ramakrishnan Srikant, *Limiting Privacy Breaches in Privacy Preserving Data Mining*, 22 PROC. ACM SIGMOD-SIGACT-SIGART SYMP. ON PRINCIPLES DATABASE SYS. 211 (2003)). This primer's presentation of the opt-out scenario versus real-world computation is influenced by Dwork, and its risk analysis is influenced by Kasiviswanathan & Smith. Cynthia Dwork, *Differential Privacy*, 33 INT'L COLLOQUIUM ON AUTOMATA, LANGUAGES & PROGRAMMING 1 (2006) [hereinafter Dwork, *Differential Privacy*]; Shiva Prasad Kasiviswanathan & Adam Smith, *On the 'Semantics' of Differential Privacy: A Bayesian Formulation*, 6 J. PRIVACY & CONFIDENTIALITY 1 (2014). For other presentations of differential privacy, see Dwork (2011) and Heffetz and Ligett (2014). Cynthia Dwork, *A Firm Foundation for Private Data Analysis*, 54 COMM. ACM 86 (2011) [hereinafter Dwork, *A Firm Foundation*]; Ori Heffetz & Katrina Ligett, *Privacy and Data-Based Research*, 28 J. ECON. PERSP. 75 (2014). For a thorough technical introduction to differential privacy, see DWORK & ROTH, *supra* note 25; Salil Vadhan, *The Complexity of Differential Privacy*, *in* TUTORIALS ON THE FOUNDATIONS OF CRYPTOGRAPHY 347 (Yehuda Lindell ed., 2017).

II. PRIVACY: A PROPERTY OF THE ANALYSIS—NOT ITS OUTPUT

This Article seeks to explain how data containing personal information can be shared in a form that ensures the privacy of the individuals in the data will be protected. The formal study of privacy in the theoretical computer science literature has yielded insights into this problem and revealed why so many traditional privacy-preserving techniques have failed to adequately protect privacy in practice. First, many traditional approaches to privacy failed to acknowledge that attackers could use information obtained from outside the system (i.e., auxiliary information) in their attempts to learn private individual information from a data release.[47] As the amount of detailed auxiliary information continues to grow and become more widely available over time, any privacy-preserving method must take auxiliary information into account in order to provide a reasonable level of privacy protection in light of any auxiliary information that an attacker may hold.[48] Furthermore, traditional approaches treated privacy as a property of the output of an analysis, whereas it is now understood that privacy should be viewed as a property of the analysis itself.[49] Any privacy-preserving method—including differential privacy—must adhere to this general principle in order to guarantee privacy protection.

The following discussion provides an intuitive explanation of these principles, beginning with a cautionary tale about the re-identification of anonymized records released by the Massachusetts Group Insurance Commission.[50]

Example 1

In the late 1990s, the Group Insurance Commission, an agency providing health insurance to Massachusetts state employees, allowed researchers to access anonymized records summarizing information about all hospital visits made by state employees. The agency anticipated that the analysis of these records would lead to recommendations for improving healthcare and controlling

47. *See* Arvind Narayanan & Vitaly Shmatikov, *Myths and Fallacies of "Personally Identifiable Information"*, 53 COMM. ACM 24, 25–26 (2010). For examples illustrating what can happen if auxiliary information is not taken into account, see Narayanan, Huey & Felten, *supra* note 30, 363–65.

48. *See* Naranayan, Huey & Felten, *supra* note 30, at 358.

49. *See id.*; Frank McSherry, *Privacy Preserving Data Analysis*, U. CAL. SANTA CRUZ, https://users.soe.ucsc.edu/~abadi/CS223_F12/mcsherry.pdf [https://perma.cc/5DJ5-KX9B] (last visited Oct. 4, 2018). For a general discussion of the advantages of formal privacy models over ad-hoc privacy techniques, see Narayanan, Huey & Felten, *supra* note 30.

50. *See Pa. Privacy Hearing*, *supra* note 26.

healthcare costs.

Massachusetts Governor William Weld reassured the public that steps would be taken to protect the privacy of patients in the data. Before releasing the records to researchers, the agency removed names, addresses, Social Security numbers, and other pieces of information that could be used to identify individuals in the records.

Viewing this as a challenge, Professor Latanya Sweeney, then a graduate student at MIT, set out to identify Governor Weld's record in the dataset. She obtained demographic information about Governor Weld, including his ZIP code and date of birth, by requesting a copy of voter registration records made available to the public for a small fee. Finding just one record in the anonymized medical claims dataset that matched Governor Weld's gender, ZIP code, and date of birth enabled her to mail the Governor a copy of his personal medical records.

As Example 1 illustrates, in many cases, a dataset that appears to be anonymous may nevertheless be used to learn sensitive information about individuals. In her demonstration, Professor Sweeney used voter registration records as auxiliary information in an attack. This re-identification demonstrates the importance of using privacy-preserving methods that are robust to auxiliary information that may be exploited by an adversary. Following Professor Sweeney's famous demonstration, a long series of attacks has been carried out against different types of data releases anonymized using a wide range of techniques and auxiliary information.[51] These attacks have shown that risks remain even if additional pieces of information, such as those that were leveraged in Professor Sweeney's attack (gender, date of birth, and ZIP code), are removed from a dataset prior to release.[52] Risks also remain when using some traditional SDL techniques, such as k-anonymity, which is satisfied for a dataset in which the identifying attributes that appear for each person are identical to those of at least $k-1$ other individuals in the dataset.[53] Research has continually demonstrated that privacy measures that treat privacy as a property of

51. *See, e.g.*, *supra* notes 26–29 and accompanying text.

52. *See, e.g.*, *supra* notes 26–29 and accompanying text.

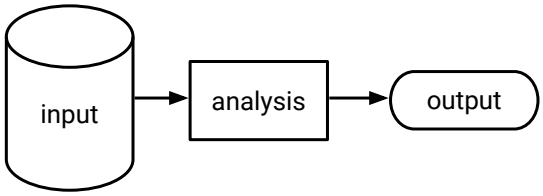
53. *See, e.g.*, Ashwin Machanavajjhala et al., *ℓ-Diversity: Privacy Beyond k-Anonymity*, 22 INT'L CONF. ON DATA ENGINEERING 24, 24 (2006) ("In this paper we show with two simple attacks that a k-anonymized dataset has some subtle, but severe privacy problems.").

the output, such as k-anonymity and other traditional statistical disclosure limitation techniques, will fail to protect privacy.

The Authors offer a brief note on terminology before proceeding. The discussions throughout this Article use the terms "analysis" and "computation" interchangeably to refer to any transformation, usually performed by a computer program, of input data into some output.

As an example, consider an analysis on data containing personal information about individuals. The analysis may be as simple as determining the average age of the individuals in the data, or it may be more complex and utilize sophisticated modeling and inference techniques. In any case, the analysis involves performing a computation on input data and outputting the result. Figure 1 illustrates this notion of an analysis.

Figure 1. An Analysis



This primer focuses, in particular, on analyses for transforming sensitive personal data into an output that can be released publicly. For example, an analysis may involve the application of techniques for aggregating or de-identifying a set of personal data in order to produce a sanitized version of the data that is safe to release. The data provider will want to ensure that publishing the output of this computation will not unintentionally leak information from the privacy-sensitive input data—but how?

A key insight from the theoretical computer science literature is that privacy is a property of the informational relationship between the input and output, not a property of the output alone.[54] The following discussion illustrates why this is the case through a series of examples.

Example 2

Anne, a staff member at a high school, would like to include statistics about student performance in a presentation. She

54. This insight follows from a series of papers demonstrating privacy breaches enabled by leakages of information resulting from decisions made by the computation. *See, e.g.*, Krishnaram Kenthapadi, Nina Mishra & Kobbi Nissim, *Denials Leak Information: Simulatable Auditing*, 79 J. COMPUTER & SYS. SCI. 1322, 1323 (2013).

considers publishing the fact that the GPA of a representative ninth-grade student is 3.5. Because the law protects certain student information held by educational institutions, she must ensure that the statistic will not inappropriately reveal student information, such as the GPA of any particular student.

One might naturally think that Anne could examine the statistic itself and determine that it is unlikely to reveal private information about an individual student. However, although the publication of this statistic might seem harmless, Anne needs to know how the statistic was computed to make that determination. For instance, if the representative ninth-grade GPA was calculated by taking the GPA of the alphabetically first student in the school, then the statistic completely reveals the GPA of that student.[55]

Example 3

Alternatively, Anne considers calculating a representative statistic based on average features of the ninth graders at the school. She takes the most common first name, the most common last name, the average age, and the average GPA for the ninth-grade class. What she produces is "John Smith, a fourteen-year-old in the ninth grade, has a 3.1 GPA." Anne includes this statistic and the method used to compute it in her presentation. In an unlikely turn of events, a new ninth-grade student named John Smith joins the class the following week.

Although the output of Anne's analysis *looks* like it reveals private information about the new ninth grader John Smith, it actually does not—because the analysis itself was not based on his student records in any way. While Anne might decide to present the statistic differently to avoid confusion, using it would not reveal private information about John. It may seem counterintuitive that releasing a "representative" GPA violates privacy (as shown by Example 2), while releasing a GPA attached to a student's name would not (as shown by Example 3). Yet these examples illustrate that the key to preserving

55. One might object that the student's GPA is not traceable back to that student unless an observer knows how the statistic was produced. However, a basic principle of modern cryptography (known as Kerckhoffs' principle) holds that a system is not secure if its security depends on its inner workings being a secret. *See* AUGUSTE KERCKHOFFS, LA CRYPTOGRAPHIE MILITAIRE [MILITARY CRYPTOGRAPHY] 8 (1883). As applied in this example, this means that it is taken as an assumption that the algorithm behind a statistical analysis is public (or could potentially be public).

privacy is the informational relationship between the private input and the public output—and not the output itself. Furthermore, not only is it necessary to examine the analysis itself to determine whether a statistic can be published while preserving privacy, but it is also sufficient. In other words, if one knows whether the process used to generate a statistic preserves privacy, the output statistic does not need to be considered at all.

III. WHAT IS THE DIFFERENTIAL PRIVACY GUARANTEE?

The previous Part illustrates why privacy should be thought of as a property of a computation—but how does one know whether a particular computation has this property?

Intuitively, a computation protects the privacy of individuals in the data if its output does not reveal any information that is specific to any individual data subject. Differential privacy formalizes this intuition as a mathematical definition.[56] Just as we can show that an integer is even by demonstrating that it is divisible by two, we can show that a computation is differentially private by proving it meets the constraints of the definition of differential privacy. In turn, if a computation can be proven to be differentially private, we can rest assured that using the computation will not unduly reveal information *specific* to any data subject.[57] Here, the term *specific* refers to information that cannot be inferred unless the individual's information is used in the analysis. For example, the information released by Anne in Example 3 is not specific to the new ninth grader John Smith because it is computed without using his information.

The following example illustrates how differential privacy formalizes this intuitive privacy requirement as a definition.

Example 4

Researchers have selected a sample of individuals across the United States to participate in a survey exploring the relationship between socioeconomic status and health outcomes. The participants were asked to complete a questionnaire covering topics concerning their residency, their finances, and their medical history.

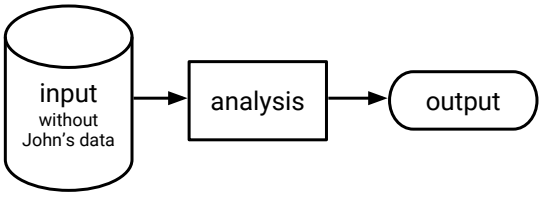
56. *See* Dwork et al., *supra* note 38, at 265–66.

57. *See id.*

> One of the participants, John, is aware that individuals have been re-identified in previous releases of de-identified data and is concerned that personal information he provides about himself, such as his medical history or annual income, could one day be revealed in de-identified data released from this study. If leaked, this information could lead to a higher life insurance premium or an adverse decision with respect to a future mortgage application.[58]

Differential privacy can be used to address John's concerns. If the researchers promise they will only share survey data after processing the data with a differentially private computation, John is guaranteed that any data the researchers release will disclose essentially nothing that is specific to him, even though he participated in the study.[59] To understand what this means, consider the thought experiment, illustrated in Figure 2 and referred to as John's opt-out scenario. In John's opt-out scenario, an analysis is performed using data about the individuals in the study, except that information about John is omitted. His privacy is protected in the sense that the outcome of the analysis does not depend on his specific information—because his information was not used in the analysis at all.

Figure 2. John's Opt-Out Scenario



John's opt-out scenario differs from the real-world scenario depicted in Figure 1, where John's information is part of the input of the analysis along with the personal information of the other study participants. In contrast to his opt-out scenario, the real-world scenario involves some potential risk to John's privacy. Some of his personal information could

58. Note that these examples are introduced for the purposes of illustrating a general category of privacy-related risks relevant to this discussion, not as a claim that life insurance and mortgage companies currently engage in this practice.

59. Intuitively, the opt-out scenario and real-world scenario are very similar, and the difference between the two scenarios is measurable and small, as described in more detail in Part IV.

be revealed by the outcome of the analysis because his information was used as input to the computation.[60]

A. Examples Illustrating What Differential Privacy Protects

Differential privacy aims to protect John's privacy in the real-world scenario in a way that mimics the privacy protection he is afforded in his opt-out scenario.[61] In other words, what can be learned about John from a differentially private computation is essentially limited to what could be learned about him from everyone else's data without his own data being included in the computation. Crucially, this same guarantee is made not only with respect to John, but also with respect to every other individual contributing her information to the analysis.

A precise description of the differential privacy guarantee requires using formal mathematical language, as well as technical concepts and reasoning that are beyond the scope of this Article. In lieu of the mathematical definition, this Article offers a few illustrative examples to discuss various aspects of differential privacy in a way designed to be intuitive and generally accessible. The scenarios in this Section illustrate the types of information disclosures that are addressed when using differential privacy.

Example 5

Alice and Bob are professors at Private University. They both have access to a database that contains personal information about students at the university, including information related to the financial aid each student receives. Because it contains personal information, access to the database is restricted. To gain access, Alice and Bob were required to demonstrate they planned to follow the university's protocols for handling personal data by undergoing confidentiality training and signing data use agreements

60. *See* Cynthia Dwork & Moni Naor, *On the Difficulties of Disclosure Prevention in Statistical Databases or the Case for Differential Privacy*, 2 J. PRIVACY & CONFIDENTIALITY 93, 95 (2008).

61. *See generally* Dwork, *Differential Privacy, supra* note 46. It is important to note that the use of differentially private analysis is *not* equivalent to the traditional use of opting out. On the privacy side, differential privacy does not require an explicit opt-out. In comparison, traditional use of opt-out may cause privacy harms by calling attention to individuals who choose to opt out. On the utility side, there is no general expectation that using differential privacy would yield the same outcomes as adopting the policy of opt-out.

proscribing their use and disclosure of personal information obtained from the database.

In March, Alice publishes an article based on the information in this database and writes that "the current freshman class at Private University is made up of 3,005 students, 202 of whom are from families earning over $350,000 per year." Alice reasons that, because she published an aggregate statistic taken from over 3,005 people, no individual's personal information will be exposed. The following month, Bob publishes a separate article containing these statistics: "201 students in Private University's freshman class of 3,004 have household incomes exceeding $350,000 per year." Neither Alice nor Bob is aware that they have both published similar information.

A clever student Eve reads both of these articles and makes an observation. From the published information, Eve concludes that between March and April one freshman withdrew from Private University and that the student's parents earn over $350,000 per year. Eve asks around and is able to determine that a student named John dropped out around the end of March. Eve then informs her classmates that John's family probably earns over $350,000 per year.

John hears about this and is upset that his former classmates learned about his family's financial status. He complains to the university, and Alice and Bob are asked to explain. In their defense, both Alice and Bob argue that they published only information that had been aggregated over a large population and does not identify any individuals.

Example 5 illustrates how, in combination, the results of multiple analyses using information about the same people may enable one to draw conclusions about individuals in the data. Alice and Bob each published information that, in isolation, seems innocuous. However, when combined, the information they published compromised John's privacy. This type of privacy breach is difficult for Alice or Bob to prevent individually, as neither knows what information others have already revealed or will reveal in future. This is referred to as the problem of composition.[62]

62. *See* Cynthia Dwork et al., *Calibrating Noise to Sensitivity in Private Data Analysis*, 7 J. PRIVACY & CONFIDENTIALITY 17, 28 (2016) (note that this article shares a title with, and is a later version of, the authors' prior paper, *supra* note 38); Srivatsava Ranjit Ganta, Shiva Prasad

Suppose, instead, that the institutional review board at Private University only allows researchers to access student records by submitting queries to a special data portal. This portal responds to every query with an answer produced by running a differentially private computation on the student records. As explained in Part IV, differentially private computations introduce a carefully tuned amount of random noise to the statistics outputted.[63] This means that the computation gives an approximate answer to every question asked through the data portal.[64] As Example 6 illustrates, the use of differential privacy prevents the privacy leakage that occurred in Example 5.

Example 6

In March, Alice queries the data portal for the number of freshmen who come from families with a household income exceeding $350,000. The portal returns the noisy count of 204, leading Alice to write in her article that "the current freshman class at Private University includes approximately 200 students from families earning over $350,000 per year." In April, Bob asks the same question and gets the noisy count of 199 students. Bob publishes in his article that "approximately 200 families in Private University's freshman class have household incomes exceeding $350,000 per year." The publication of these noisy figures prevents Eve from concluding that one student, with a household income greater than $350,000, withdrew from the university in March. The risk that John's personal information could be uncovered based on these publications is thereby reduced.

Example 6 hints at one of the most important properties of differential privacy—it is robust under composition.[65] If multiple analyses are performed on data describing the same set of individuals, then, as long as each of the analyses satisfies differential privacy, it is guaranteed that all of the information released, when taken together, will still be differentially private.[66] Notice how this example is

Kasiviswanathan & Adam Smith, *Composition Attacks and Auxiliary Information in Data Privacy*, 14 PROC. ACM SIGKDD INT'L CONF. ON KNOWLEDGE, DISCOVERY & DATA MINING 265, 265–66 (2008).

63. *See infra* Part IV.

64. *See infra* Part IV.

65. *See* Vadhan, *supra* note 46, at 348–49.

66. *See id.* at 349, 361.

markedly different from Example 5, in which Alice and Bob do not use differentially private analyses and inadvertently release two statistics that, when combined, lead to the full disclosure of John's personal information. The use of differential privacy rules out the possibility of such a complete breach of privacy. This is because differential privacy enables one to measure and bound the cumulative privacy risk from multiple analyses of information about the same individuals.[67]

It is important to note, however, that every analysis, regardless of whether it is differentially private or not, results in some leakage of information about the individuals whose information is being analyzed. This is a well-established principle within the statistical community, as evidenced by a 2005 report that concluded "[t]he release of statistical data inevitably reveals some information about individual data subjects."[68] Furthermore, this leakage accumulates with each analysis, potentially to a point where an attacker may infer the underlying data.[69] This is true for every release of data, including releases of aggregate statistics.[70] In particular, releasing too many aggregate statistics too accurately inherently leads to severe privacy loss.[71] For this reason, there is a limit to how many analyses can be performed on a specific dataset while providing an acceptable guarantee of privacy.[72] This is why it is critical to measure privacy loss and to understand quantitatively how risk accumulates across successive analyses, as Sections IV.E and VI.A describe below.

B. Examples Illustrating What Differential Privacy Does Not Protect

The following examples illustrate the types of information disclosures differential privacy does not seek to address.

Example 7

Suppose Ellen is a friend of John's and knows some of his habits, such as that he regularly consumes several glasses of red wine with

67. *See id.*

68. *See* Fed. Comm. on Statistical Methodology, *supra* note 19, at 3.

69. *See, e.g.*, Dinur & Nissim, *supra* note 30, at 203; Cynthia Dwork et al., *Exposed! A Survey of Attacks on Private Data*, 4 ANN. REV. STAT. & ITS APPLICATION 61, 64 (2016); Nils Homer et al., *Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays*, 4 PLoS Genetics e1000167, at 6, 9 (2008), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2516199/pdf/pgen.1000167.pdf [https://perma.cc/7873-CG6L]; Fed. Comm. on Statistical Methodology, *supra* note 19, at 3.

70. *See* sources cited *supra* note 69.

71. *See* sources cited *supra* note 69.

72. *See* sources cited *supra* note 69.

dinner. Ellen learns that John took part in a large research study, and that this study found a positive correlation between drinking red wine and the likelihood of developing a certain type of cancer. She might therefore conclude, based on the results of this study and her prior knowledge of John's drinking habits, that he has a heightened risk of developing cancer.

It may seem at first that the publication of the results from the research study enabled a privacy breach by Ellen. After all, learning about the study's findings helped her infer new information about John that he himself may be unaware of (i.e., his elevated cancer risk). However, notice that Ellen would be able to infer this information about John even if John had not participated in the medical study (i.e., it is a risk that exists in both John's opt-out scenario and the real-world scenario).[73] Risks of this nature apply to everyone, regardless of whether they shared personal data through the study or not. Consider another example:

Example 8

Ellen knows that her friend John is a public school teacher with five years of experience and that he is about to start a job in a new school district. She later comes across a local news article about a teachers' union dispute, which includes salary figures for the public school teachers in John's new school district. Ellen is able to approximately determine John's salary at his new job, based on the district's average salary for a teacher with five years of experience.

Note that, as in the previous example, Ellen can determine information about John (i.e., his new salary) from the published information, even though the published information was not based on John's information. In both examples, John could be adversely affected by the discovery of the results of an analysis, even in his opt-out scenario. In both John's opt-out scenario and in a differentially private real-world scenario, it is therefore not guaranteed that no information about John can be revealed. The use of differential privacy limits the revelation of information *specific* to John.

73. Ellen's inference would rely on factors such as the size of the study sample, whether the sampling was performed at random, and whether John comes from the same population as the sample, among others.

These examples suggest, more generally, that any useful analysis carries a risk of revealing some information about individuals. One might observe, however, that such risks are largely unavoidable. In a world in which data about individuals are collected, analyzed, and published, John cannot expect better privacy protection than is offered by his opt-out scenario because he has no ability to prevent others from participating in a research study or appearing in public records.

Moreover, the types of information disclosures enabled in John's opt-out scenario often result in individual and societal benefits. For example, the discovery of a causal relationship between red wine consumption and elevated cancer risk can lead to new public health recommendations, support future scientific research, and inform John about possible changes he could make in his habits that would likely have positive effects on his health. Similarly, the publication of public school teacher salaries may be seen as playing a critical role in transparency and public policy, as it can help communities make informed decisions regarding appropriate salaries for their public employees.

IV. HOW DOES DIFFERENTIAL PRIVACY LIMIT PRIVACY LOSS?

The previous Part explains that the only things that can be learned about a data subject from a differentially private data release are essentially what could have been learned if the analysis had been performed without that individual's data.

How do differentially private analyses achieve this goal? And what is meant by "essentially" when stating that the only things that can be learned about a data subject are essentially those things that could be learned without the data subject's information? The answers to these two questions are related. Differentially private analyses protect the privacy of individual data subjects by introducing carefully tuned random noise when producing statistics.[74] Differentially private analyses are also allowed to leak *some* small amount of information specific to individual data subjects.[75] A privacy parameter controls exactly how much information can be leaked and, relatedly, how much random noise is introduced during the differentially private computation.[76]

74. *See* Dwork et al., *supra* note 38, at 265–66.
75. *See id.* at 267.
76. Dwork et al., *supra* note 62, at 18.

A. Differential Privacy and Randomness

Example 6 shows that differentially private analyses introduce random noise to the statistics they produce. Intuitively, this noise masks the differences between the real-world computation and the opt-out scenario of each individual in the dataset. This means that the outcome of a differentially private analysis is not exact, but rather an approximation. In addition, a differentially private analysis may, if performed twice on the same dataset, return different results because it intentionally introduces random noise. Therefore, analyses performed with differential privacy differ from standard statistical analyses, such as the calculation of averages, medians, and linear regression equations, in which one gets the same answer when a computation is repeated twice on the same dataset.

Example 9

Consider a differentially private analysis that computes the number of students in a sample with a GPA of at least 3.0. Say that there are 10,000 students in the sample, and exactly 5,603 of them have a GPA of at least 3.0. An analysis that added no random noise would report that 5,603 students had a GPA of at least 3.0.

A differentially private analysis, however, introduces random noise to protect the privacy of the data subjects. For instance, a differentially private analysis might report an answer of 5,521 when run on the student data; when run a second time on the same data, it might report an answer of 5,586.[77]

Although a differentially private analysis might produce many different answers given the same dataset, it is usually possible to calculate accuracy bounds for the analysis measuring how much an output of the analysis is expected to differ from the noiseless answer.[78] Section VI.B discusses how the random noise introduced by a differentially private analysis affects statistical accuracy. Appendix A.1

77. Note that, if an analyst is allowed to repeat this computation multiple times, she could average out the noise and get the exact answer. The number of allowable repetitions is limited by an overall privacy budget. *See infra* Section VI.A.

78. *See, e.g.*, DWORK & ROTH, *supra* note 25, at 22; Prashanth Mohan et al., *GUPT: Privacy Preserving Data Analysis Made Easy*, 2012 PROC. ACM SIGMOD INT'L CONF. ON MGMT. DATA 349, 349; Vadhan, *supra* note 46, at 366–67; Marco Gaboardi et al., *PSI (ψ): A Private Data Sharing Interface* 15 (ArXiv, Working Paper No. 1609.04340, 2018), https://arxiv.org/pdf/1609.04340.pdf [https://perma.cc/PXC4-6CEL].

provides more information about the role randomness plays in the construction of differentially private analyses.

B. The Privacy Loss Parameter

An essential component of a differentially private computation is the privacy loss parameter, which determines how well each individual's information needs to be hidden and, consequently, how much noise needs to be introduced.[79] It can be thought of as a tuning knob for balancing privacy and accuracy. Each differentially private analysis can be tuned to provide more or less privacy—resulting in less or more accuracy, respectively—by changing the value of this parameter. The parameter can be thought of as limiting how much a differentially private computation is allowed to deviate from the opt-out scenario of each individual in the data.

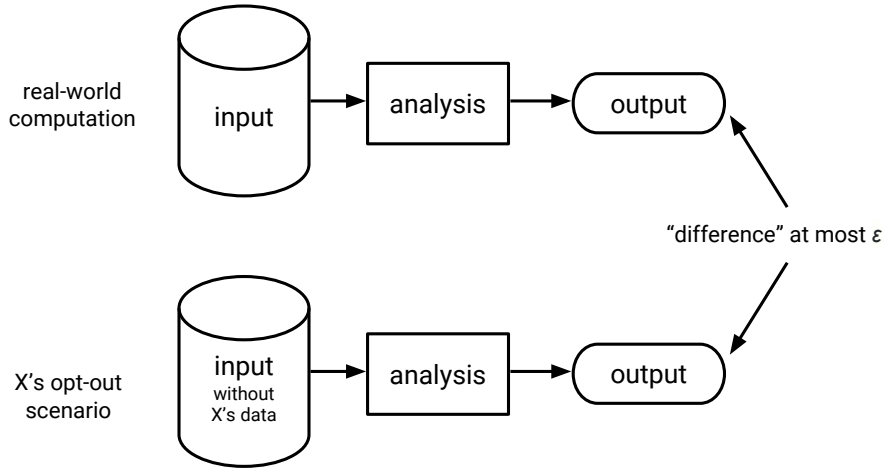
Consider the opt-out scenario for a certain computation, such as estimating the number of HIV-positive individuals in a surveyed population. Ideally, this estimate should remain exactly the same whether or not a single individual, such as John discussed above, is included in the survey. However, as described above, ensuring that the estimate is *exactly* the same would require the total exclusion of John's information from the real-world analysis. It would also require excluding the information of other individuals (e.g., that of Gertrude, Peter, and so forth) in order to provide perfect privacy protection for them as well. Continuing this line of argument, one can conclude that the personal information of every single surveyed individual must be removed in order to satisfy each individual's opt-out scenario. Thus, the analysis cannot rely on any person's information and is completely useless.

To avoid this dilemma, differential privacy requires only that the output of the analysis remain approximately the same, whether John participates in the survey or not. That is, differential privacy allows for a deviation between the output of the real-world analysis and that of each individual's opt-out scenario. A parameter quantifies and limits the extent of the deviation between the opt-out and real-world scenarios.[80] As Figure 3 illustrates below, this parameter is usually denoted by the Greek letter ε (epsilon) and referred to as the privacy parameter or, more accurately, the privacy loss parameter.[81] The parameter ε measures the effect of each individual's information on the

79. *See* DWORK & ROTH, *supra* note 25, at 6.
80. *Id.*
81. *See id.*

output of the analysis. It can also be viewed as a measure of the additional privacy risk an individual could incur beyond the risk incurred in the opt-out scenario. Note that Figure 3 replaces John with an arbitrary individual *X* to emphasize that the differential privacy guarantee is made simultaneously to all individuals in the sample—not just John.

Figure 3. Differential Privacy



Moreover, it can be shown that the deviation between the real-world and opt-out scenarios cannot be increased by any further processing of the output of a differentially private analysis. Hence, the guarantees of differential privacy, described below, hold regardless of how an attacker may try to manipulate the output. In this sense, differential privacy is robust to a wide range of potential privacy attacks, including attacks that are unknown at the time of deployment.[82]

Choosing a value for ε can be thought of as setting the desired level of privacy protection. This choice also affects the utility or accuracy that can be obtained from the analysis.[83] A smaller value of ε results in a smaller deviation between the real-world analysis and each opt-out scenario and is therefore associated with stronger privacy

82. The property that differential privacy is preserved under arbitrary further processing is referred to as (resilience to) post-processing. *See* DWORK & ROTH, *supra* note 25, at 19.

83. *See id.* For an illustration of how the choice of epsilon can affect accuracy, see *infra* Figure 4.

protection but less accuracy.[84] For example, when ε is set to zero, the real-world differentially private analysis mimics the opt-out scenario of each individual perfectly and simultaneously. However, an analysis that perfectly mimics the opt-out scenario of each individual would require ignoring all information from the input and, accordingly, could not provide any meaningful output. Yet, when ε is set to a small number such as 0.1, the deviation between the real-world computation and each individual's opt-out scenario will be small, providing strong privacy protection, while also enabling an analyst to derive useful statistics based on the data.

Accepted guidelines for choosing ε have not yet been developed.[85] The increasing use of differential privacy in real-life applications will likely shed light on how to reach a reasonable compromise between privacy and accuracy, and the accumulated evidence from these real-world decisions will likely contribute to the development of future guidelines.[86] As discussed in Section IV.D, the Authors of this Article recommend that, when possible, ε be set to a small number, such as a

84. *See infra* Figure 4.

85. *See* JOHN M. ABOWD & IAN M. SCHMUTTE, REVISITING THE ECONOMICS OF PRIVACY: POPULATION STATISTICS AND CONFIDENTIALITY PROTECTION AS PUBLIC GOODS 1 (2015), https://digitalcommons.ilr.cornell.edu/cgi/viewcontent.cgi?article=1036&context=ldi [https://perma.cc/8B8Q-LCFA]; GARFINKEL, *supra* note 9, at 54; Justin Hsu et al., *Differential Privacy: An Economic Method for Choosing Epsilon*, 27 IEEE COMPUTER SECURITY FOUND. SYMP. 398, 398 (2014). *See generally* John M. Abowd & Ian M. Schmutte, *An Economic Analysis of Privacy Protection and Statistical Accuracy as Social Choices*, AM. ECON. REV. (forthcoming).

86. Setting the primary loss parameter ε is a policy decision to be informed by normative and technical considerations. Companies and governments experimenting with practical implementations of differential privacy have selected various values for ε. Some of these implementations have adopted values of ε exceeding 1 due to the difficulty of meeting utility requirements using lower values of ε. To date, these choices of ε have not led to known vulnerabilities. For example, the US Census Bureau reportedly chose a value of $\varepsilon = 8.9$ for OnTheMap—a public interface which allows users to explore American commuting patterns using a variant of differential privacy. *See* John M. Abowd, Assoc. Dir. for Research and Methodology, US Census Bureau, The Challenge of Scientific Reproducibility and Privacy Protection for Statistical Agencies, Presentation for the Census Scientific Advisory Committee 12 (Sept. 15, 2016), https://www2.census.gov/cac/sac/meetings/2016-09/2016-abowd.pdf [https://perma.cc/4CXN-C257]. As another example, researchers have determined that Apple's differential private data collection in macOS 10.12 and iOS 10 likely uses values of ε as high as 6 and 14, respectively. *See* Jun Tang et al., *Privacy Loss in Apple's Implementation of Differential Privacy on MacOS 10.12* (ArXiv, Working Paper No. 1709.02753, 2017), https://arxiv.org/pdf/1709.02753.pdf [https://perma.cc/V4QE-QJ49]. Although differential privacy is an emerging concept and has been deployed in limited applications to date, best practices may emerge over time as values for ε are selected for implementations of differential privacy in a wide range of settings. With this in mind, researchers have proposed that a registry be created to document details of differential privacy implementations, including the value of ε chosen and the factors that led to its selection. *See* NAT'L ACAD. OF SCIS., ENG'G & MED., FEDERAL STATISTICS, MULTIPLE DATA SOURCES, AND PRIVACY PROTECTION: NEXT STEPS 107 (Robert M. Groves & Brian A. Harris-Kojetin eds., 2017) (citing Cynthia Dwork & Dierdre Mulligan, Differential Privacy in Practice: Expose Your Epsilons! (June 5, 2014) (unpublished manuscript)), http://nap.edu/24893 [https://perma.cc/5YKH-QQBG].

value less than 1.[87] As Figure 3 illustrates, the maximum deviation between the opt-out scenario and the real-world computation should hold simultaneously for each individual X whose information is included in the input.

C. Bounding Risk

The previous Section discusses how the privacy loss parameter limits the deviation between the real-world computation and each data subject's opt-out scenario. However, it might not be clear how this abstract guarantee relates to the privacy concerns individuals face in the real world. To help ground the concept, this Section discusses a practical interpretation of the privacy loss parameter. It describes how the parameter can be understood as a bound on the financial risk incurred by an individual participating in a research study.

Any useful analysis carries the risk that it will reveal information about the individuals in the data.[88] An individual whose information is used in an analysis may be concerned that a potential leakage of her personal information could result in reputational, financial, or other costs. Examples 10 and 11 below introduce a scenario in which an individual participating in a research study worries that an analysis on the data collected in the research study may leak information that could lead to a substantial increase in her life insurance premium. Example 12 illustrates that, while differential privacy necessarily cannot fully eliminate this risk, it can guarantee that the risk will be limited by quantitative bounds that depend on ε.[89]

Example 10

Gertrude, a sixty-five-year-old woman, is considering whether to participate in a medical research study. While she can envision many potential personal and societal benefits resulting in part from her participation in the study, she is concerned that the personal information she discloses over the course of the study could lead to an increase in her life insurance premium in the future.

For example, Gertrude is concerned that the tests she would undergo as part of the research study would reveal that she is predisposed to suffer a stroke and is significantly more likely to die

87. *See* discussion following Table 1.
88. *See supra* Part III.
89. *See* Dwork et al., *supra* note 38, at 266–67.

in the coming year than the average person of her age and gender. If such information related to Gertrude's increased risk of morbidity and mortality is discovered by her life insurance company, it will likely increase the premium for her annual renewable term policy substantially.

Before she opts to participate in the study, Gertrude wishes to be assured that privacy measures are in place to ensure that her participation will have, at most, a limited effect on her life insurance premium.

1. A Baseline: Gertrude's Opt-Out Scenario

It is important to note that Gertrude's life insurance company may raise her premium based on something it learns from the medical research study, even if Gertrude does not herself participate in the study. The following example is provided to illustrate such a scenario.[90]

Example 11

Gertrude holds a $100,000 life insurance policy. Her life insurance company has set her annual premium at $1,000, i.e., 1% of $100,000, based on actuarial tables showing that someone of Gertrude's age and gender has a 1% chance of dying in the next year.

Suppose Gertrude opts out of participating in the medical research study. Regardless, the study reveals that coffee drinkers are more likely to suffer a stroke than non-coffee drinkers. Gertrude's life insurance company may update its assessment and conclude that, as a sixty-five-year-old woman who drinks coffee, Gertrude has a 2% chance of dying in the next year. The company decides to increase Gertrude's annual premium from $1,000 to $2,000 based on the findings of the study.[91]

90. Figures in this example are based on data from *Actuarial Life Table: Period Life Table, 2015*, SOC. SECURITY ADMIN., http://www.ssa.gov/oact/STATS/table4c6.html [https://perma.cc/7ZPH-GE7N] (last visited Sept. 22, 2018).

91. Note that there may be legal, policy, or other reasons why a company would not raise Gertrude's insurance premium based on the outcome of this study. Also, this is not a claim that insurance companies engage in this practice. Example 11 is introduced for the purposes of illustrating a general category of privacy-related risks relevant to this discussion. This example assumes that the insurance company updates its belief about Gertrude's chances of dying next year based on the outcome of this study using a Bayesian analysis. Furthermore, it assumes that Gertrude's premium is then updated in proportion to this change in belief. Differential privacy also

In this example, the results of the study led to an increase in Gertrude's life insurance premium, even though she did not contribute any personal information to the study. A potential increase of this nature is unavoidable to Gertrude in this scenario because she cannot prevent other people from participating in the study. This example illustrates that Gertrude can experience a financial loss even in her opt-out scenario. Because, as presented in this example, Gertrude cannot avoid this type of risk on her own,[92] in the following discussion this opt-out scenario will serve as a baseline for measuring potential increases in her privacy risk above this threshold.

2. Reasoning About Gertrude's Risk

Next consider the increase in risk, relative to Gertrude's opt-out scenario, that is due to her participation in the study.

Example 12

Suppose Gertrude decides to participate in the research study. Based on the results of medical tests performed on Gertrude over the course of the study, the researchers conclude that Gertrude has a 50% chance of dying from a stroke in the next year. If the data from the study were to be made available to Gertrude's insurance company, it might decide to increase her insurance premium to $50,000 in light of this discovery.

Fortunately for Gertrude, this does not happen. Rather than releasing the full dataset from the study, the researchers release only a differentially private summary of the data they collected. Differential privacy guarantees that, if the researchers use a value of $\varepsilon = 0.01$, then the insurance company's estimate of the probability that Gertrude will die in the next year can increase from the opt-out scenario's estimate of 2% to at most

$$2\% \cdot (1 + 0.01) = 2.02\%.$$

allows one to reason (in a different manner) about a more general case where no assumptions are made regarding how the insurance company updates Gertrude's premium, but that analysis is omitted from this discussion for simplicity.

92. Although Gertrude, acting as an individual, cannot avoid this risk, society or groups of individuals may collectively act to avoid such a risk. For example, the researchers could be prohibited from running the study, or the data subjects could collectively decide not to participate. Therefore, the use of differential privacy does not completely eliminate the need to make policy decisions regarding the value of allowing data collection and analysis in the first place.

Thus Gertrude's insurance premium can increase from $2,000 to, at most, $2,020. Gertrude's first-year cost of participating in the research study, in terms of a potential increase in her insurance premium, is at most $20.

Note that this does not mean that the insurance company's estimate of the probability that Gertrude will die in the next year will necessarily increase as a result of her participation in the study, nor that if the estimate increases it must increase to 2.02%. What the analysis shows is that if the estimate were to increase it would not exceed 2.02%.

In this example, Gertrude is aware of the fact that the study could indicate that her risk of dying in the next year exceeds 1%. She happens to believe, however, that the study will not indicate more than a 2% risk of dying in the next year, in which case the potential cost to her of participating in the research will be at most $20. Based on her belief, Gertrude may decide that she considers the potential cost of $20 to be too high and that she cannot afford to participate with this value of ε and this level of risk. Alternatively, she may decide that it is worthwhile. Perhaps she is paid more than $20 to participate in the study, or the information she learns from the study is worth more than $20 to her. The key point is that differential privacy allows Gertrude to make a more informed decision based on the worst-case cost of her participation in the study.

It is worth noting that, should Gertrude decide to participate in the study, her risk might increase—even if her insurance company is not aware of her participation. Gertrude might actually have a higher chance of dying in the next year, and that could affect the study results. In turn, her insurance company might decide to raise her premium because she fits the profile of the studied population—even if it does not believe her data were included in the study. Differential privacy guarantees that, even if the insurance company knows that Gertrude *did* participate in the study—it can only make inferences about her that it could have essentially made if she had not participated in the study.

D. A General Framework for Reasoning About Privacy Risk

Gertrude's scenario illustrates how differential privacy is a general framework for reasoning about the increased risk that is incurred when an individual's information is included in a data analysis. Differential privacy guarantees that an individual will be exposed to essentially the same privacy risk, whether or not her data

are included in a differentially private analysis.[93] In this context, one can think of the privacy risk associated with a release of the output of a data analysis as the potential harm that an individual might incur because of a belief that an observer forms based on that data release.

In particular, when ε is set to a small value, an observer's posterior belief can change—relative to the case where the data subject is not included in the data set—by a factor of at most approximately $1 + \varepsilon$ based on a differentially private data release.[94] For example, if ε is set to 0.01, then the privacy risk to an individual resulting from participation in a differentially private computation grows by at most a multiplicative factor of 1.01.

As Examples 11 and 12 illustrate, there is a risk to Gertrude that the insurance company will see the study results, update its beliefs about the mortality of Gertrude, and charge her a higher premium. If the insurance company infers from the study results that Gertrude has probability p of dying in the next year and her insurance policy is valued at $100,000, her premium will increase to $p \times \$100{,}000$. This risk exists, even if Gertrude does not participate in the study. Recall how, in Example 11, the insurance company's belief that Gertrude will die in the next year doubles from 1% to 2%, increasing her premium from $1,000 to $2,000, based on general information learned from the individuals who did participate. Recall also that if Gertrude does decide to participate in the study (as in Example 12), differential privacy limits the change in this risk relative to her opt-out scenario. In financial terms, her risk increases by at most $20, since the insurance company's beliefs about her probability of death change from 2% to at most $2\% \cdot (1 + \varepsilon) = 2.02\%$, where $\varepsilon = 0.01$.

Note that the above calculation requires certain information that may be difficult to determine in the real world. In particular, the 2% baseline in Gertrude's opt-out scenario (i.e., Gertrude's insurer's belief about her chance of dying in the next year) is dependent on the results from the medical research study, which Gertrude does not know at the time she makes her decision whether to participate. Fortunately, differential privacy provides guarantees relative to every baseline risk.[95]

---

93. *See* Dwork et al., *supra* note 62, at 19; Dwork & Naor, *supra* note 60, at 103.

94. In general, the guarantee made by differential privacy is that the probabilities differ by at most a factor of $e^{\pm\varepsilon}$, which is approximately $1 \pm \varepsilon$ when ε is small. *See* Shiva Prasad Kasiviswanathan & Adam Smith, *On the 'Semantics' of Differential Privacy: A Bayesian Formulation*, 6 J. PRIVACY & CONFIDENTIALITY 1 (2014).

95. *See infra* Table 1 and accompanying text.

Example 13

Say that, without her participation, the study results would lead the insurance company to believe that Gertrude has a 3% chance of dying in the next year (instead of the 2% chance hypothesized earlier). This means that Gertrude's insurance premium would increase to $3,000. Differential privacy guarantees that, if Gertrude had instead decided to participate in the study, the insurer's estimate for Gertrude's mortality would have been at most $3\% \cdot (1 + \varepsilon) = 3.03\%$ (assuming an ε of 0.01), which means that her premium would not increase beyond $3,030.

Calculations like those used in the analysis of Gertrude's privacy risk can be performed by referring to Table 1. For example, the value of ε used in the research study Gertrude considered participating in was 0.01, and the baseline privacy risk in her opt-out scenario was 2%. As shown in Table 1, these values correspond to a worst-case privacy risk of 2.02% in her real-world scenario. Notice also how the calculation of risk would change with different values. For example, if the privacy risk in Gertrude's opt-out scenario were 5% rather than 2% and the value of ε remained the same, then the worst-case privacy risk in her real-world scenario would be 5.05%.

Table 1. Maximal Difference Between Posterior Beliefs in Gertrude's Opt-Out and Real-World Scenarios

The notation $A(x')$ refers to the application of the analysis A on the dataset x', which does not include Gertrude's information. As this table shows, the use of differential privacy provides a quantitative bound on how much one can learn about an individual from a computation.[96]

posterior belief given $A(x')$ in %	value of ε					
	0.01	0.05	0.1	0.2	0.5	1
0	0	0	0	0	0	0
1	1.01	1.05	1.1	1.22	1.64	2.67
2	2.02	2.1	2.21	2.43	3.26	5.26
5	5.05	5.24	5.5	6.04	7.98	12.52
10	10.09	10.46	10.94	11.95	15.48	23.2
25	25.19	25.95	26.92	28.93	35.47	47.54
50	50.25	51.25	52.5	54.98	62.25	73.11
75	75.19	75.93	76.83	78.56	83.18	89.08
90	90.09	90.44	90.86	91.66	93.69	96.07
95	95.05	95.23	95.45	95.87	96.91	98.1
98	98.02	98.1	98.19	98.36	98.78	99.25
99	99.01	99.05	99.09	99.18	99.39	99.63
100	100	100	100	100	100	100
	maximum posterior belief given $A(x)$ in %					

The fact that the differential privacy guarantee applies to every privacy risk means that Gertrude can know for certain how participating in the study might increase her risks relative to opting out, even if she does not know a priori all the privacy risks posed by the data release. This enables Gertrude to make a more informed decision about whether to take part in the study. For instance, perhaps with the help of the researcher obtaining her informed consent, Gertrude can use this framework to better understand how the additional risk she may incur by participating in the study is bounded. By considering the bound with respect to a range of possible baseline risk values, she may

96. For p, the posterior belief given $A(x')$, and privacy parameter ε, the bound on the posterior belief given $A(x)$ is $\frac{p}{p+e^{-\varepsilon}(1-p)}$. For small ε and p, this expression can be approximated as $p(1+\varepsilon)$. These formulas are derived from the definition of differential privacy. *See* Kobbi Nissim, Claudio Orlandi & Rann Smorodinsky, *Privacy-Aware Mechanism Design*, 13 PROC. ACM CONF. ON ELECTRONIC COM. 774, 775–89 (2012).

decide whether she is comfortable with taking on the risks entailed by these different scenarios.

Table 1 demonstrates how significant changes in posterior belief compared to the opt-out baseline can be for different values of ε. Notice how, at $\varepsilon = 1$, a belief that Gertrude has a certain condition with 1% probability in the opt-out scenario would become 2.67%, which is quite a large factor increase (more than double), and a 50% belief would become nearly a 75% belief (also a very significant change). For $\varepsilon = 0.2$ and $\varepsilon = 0.5$, the changes start to become more modest, but could still be considered too large, depending on how sensitive the data are. For $\varepsilon = 0.1$ and below, the changes in beliefs may be deemed small enough for most applications.

Also note that the entries in Table 1 are the worst-case bounds that are guaranteed by a given setting of ε. An adversary's actual posterior beliefs given *A(x)* may be smaller in a given practical application, depending on the distribution of the data, the specific differentially private algorithms used, and the adversary's prior beliefs and auxiliary information. That is, in a real-world application, a particular choice of ε may turn out to be safer than Table 1 indicates, but it can be difficult to quantify how much safer.

The exact choice of ε is a policy decision that should depend on the sensitivity of the data, with whom the output will be shared, the intended data analysts' accuracy requirements, and other technical and normative factors. Table 1 and explanations interpreting it, such as the examples provided in this Section, can help provide the kind of information needed to make such a policy decision.

E. Composition

Privacy risk accumulates with multiple analyses on an individual's data, and this is true whether or not any privacy-preserving technique is applied.[97] One of the most powerful features of differential privacy is its robustness under composition.[98] One can reason about—and bound—the privacy risk that accumulates when multiple differentially private computations are performed on an individual's data.[99]

97. *See* DWORK & ROTH, *supra* note 25, at 5. Note that this observation is not unique to differentially private analyses. It is true for *any* use of information, and, therefore, for any approach to preserving privacy. However, the fact that the cumulative privacy risk from multiple analyses can be bounded is a distinguishing property of differential privacy.

98. *See* sources cited *supra* note 62.

99. *See* sources cited *supra* note 62.

The parameter ε quantifies how privacy risk accumulates across multiple differentially private analyses. Imagine that two differentially private computations are performed on datasets about the same individuals. If the first computation uses a parameter of ε_1 and the second uses a parameter of ε_2, then the cumulative privacy risk resulting from these computations is no greater than the risk associated with an aggregate parameter of $\varepsilon_1 + \varepsilon_2$.[100] In other words, the privacy risk from running the two analyses is bounded by the privacy risk from running a single differentially private analysis with a parameter of $\varepsilon_1 + \varepsilon_2$.

Example 14

Suppose that Gertrude decides to opt into the medical study because it is about heart disease, an area of research she considers critically important. The study leads to a published research paper, which includes results from the study produced by a differentially private analysis with a parameter of $\varepsilon_1 = 0.01$. A few months later, the researchers decide that they want to use the same study data for another paper. This second paper would explore a hypothesis about acid reflux disease, and would require calculating new statistics based on the original study data. Like the analysis results in the first paper, these statistics would be computed using differential privacy, but this time with a parameter of $\varepsilon_2 = 0.02$.

Because she only consented to her data being used in research about heart disease, the researchers must obtain Gertrude's permission to reuse her data for the paper on acid reflux disease. Gertrude is concerned that her insurance company could compare the results from both papers and learn something negative about Gertrude's life expectancy and drastically raise her insurance premium. She is not particularly interested in participating in a research study about acid reflux disease and is concerned the risks of participation might outweigh the benefits to her.

Because the statistics from each study are produced using differentially private analyses, Gertrude can precisely bound the privacy risk that would result from contributing her data to the second study. The combined analyses can be thought of as a single analysis with a privacy loss parameter of

100. *See* Dwork et al., *supra* note 62, at 28.

$$\varepsilon_1 + \varepsilon_2 = 0.01 + 0.02 = 0.03.$$

Say that, without her participation in either study, the insurance company would believe that Gertrude has a 2% chance of dying in the next year, leading to a premium of $2,000. If Gertrude participates in both studies, the insurance company's estimate of Gertrude's mortality would increase to at most

$$2\% \cdot (1 + 0.03) = 2.06\%.$$

This corresponds to a premium increase of $60 over the premium that Gertrude would pay if she had not participated in either study.

This means that, while it cannot get around the fundamental law that privacy risk increases when multiple analyses are performed on the same individual's data, differential privacy guarantees that privacy risk accumulates in a bounded way.[101] Despite the accumulation of risk, two differentially private analyses cannot be combined in a way that leads to a privacy breach that is disproportionate to the privacy risk associated with each analysis in isolation. To the Authors' knowledge, differential privacy is currently the only known framework with quantifiable guarantees with respect to how risk accumulates across multiple analyses.

V. WHAT TYPES OF ANALYSES ARE PERFORMED WITH DIFFERENTIAL PRIVACY?

A large number of analyses can be performed with differential privacy guarantees. Differentially private algorithms are known to exist for a wide range of statistical analyses such as count queries, histograms, cumulative distribution functions, and linear regression; techniques used in statistics and machine learning such as clustering and classification; and statistical disclosure limitation techniques like synthetic data generation, among many others.

For the purposes of illustrating that broad classes of analyses can be performed using differential privacy, the discussion in this Part provides a brief overview of each of these types of analyses and how they can be performed with differential privacy guarantees.[102]

101. *See id.* at 28–29.

102. The discussion in this Part provides only a brief introduction to a number of statistical and machine learning concepts. For a more detailed introduction to these concepts, see, for example, JOSEPH K. BLITZSTEIN & JESSICA HWANG, INTRODUCTION TO PROBABILITY (2015);

- **Count queries:** The most basic statistical tool, a count query, returns an estimate of the number of individual records in the data satisfying a specific predicate.[103] For example, a count query could be used to return the number of records corresponding to HIV-positive individuals in a sample. Differentially private answers to count queries can be obtained through the addition of random noise, as demonstrated in the detailed example found in Appendix A.1.
- **Histograms:** A histogram contains the counts of data points as they are classified into disjoint categories.[104] For example, in the case of numerical data, a histogram shows how data are classified within a series of consecutive non-overlapping intervals. A **contingency table (or cross tabulation)** is a special form of histogram representing the interrelation between two or more variables.[105] The categories of a contingency table are defined as conjunctions of attribute variables, such as the number of individuals in a dataset that are both college-educated *and* earn less than $50,000 per year.[106] Differentially private histograms and contingency tables provide noisy counts for the data classified in each category.[107]
- **Cumulative distribution function (CDF):** For data over an ordered domain, such as age (where the domain is integers, say, in the range of 0, 1, 2, …, 100), or annual income (where the domain is real numbers, say, in the range of $0.00 – $1,000,000.00), a cumulative distribution function depicts for every domain value x an estimate of the number of data points with a value up to x.[108] A CDF can be used for computing the median of the data points

GARETH JAMES ET AL., AN INTRODUCTION TO STATISTICAL LEARNING WITH APPLICATIONS IN R 127–75 (2013).

103. *See* Mark Bun, A Teaser for Differential Privacy 1 (Dec. 8, 2017) (unpublished manuscript), https://www.cs.princeton.edu/~smattw/Teaching/521fa17lec22.pdf [https://perma.cc/L54G-BKUW].

104. *See* JOHN M. CHAMBERS ET AL., GRAPHICAL METHODS FOR DATA ANALYSIS 24–26 (1983).

105. *See* YVONNE M. BISHOP, STEPHEN E. FIENBERG & PAUL W. HOLLAND, DISCRETE MULTIVARIATE ANALYSIS: THEORY AND PRACTICE 9–13 (1975).

106. *See id.*

107. *See, e.g.*, Dwork et al., *supra* note 38, at 273.

108. *See* JAMES E. GENTLE, COMPUTATIONAL STATISTICS 29–30 (2009).

(the value x for which half the data points have value up to x) and the interquartile range, among other statistics.[109] A differentially private estimate of the CDF introduces noise that needs to be taken into account when the median or interquartile range is computed from the estimated CDF.[110]

- **Linear regression:** Social scientists are often interested in modeling how a dependent variable varies as a function of one or more explanatory variables. For instance, a researcher may seek to understand how a person's health depends on her education and income. In linear regression, an underlying linear model is assumed, and the goal of the computation is to fit a linear model to the data that minimizes a measure of "risk" (or "cost"), usually the sum of squared errors.[111] Using linear regression, social scientists can learn to what extent a linear model explains their data, and which of the explanatory variables correlates best with the dependent variable.[112] Differentially private implementations of linear regression introduce noise in its computation.[113]
- **Clustering:** Clustering is a data analysis technique that involves grouping data points into clusters, so that points in the same cluster are more similar to each other than to points in other clusters.[114] Data scientists often use clustering as an exploratory tool to gain insight into their data and identify the data's important subclasses.[115] Researchers are developing a variety of differentially private clustering algorithms,[116] and such tools are likely

109. *See id.* at 62–63, 330.

110. For a more in-depth discussion of differential privacy and CDFs, see Daniel Muise & Kobbi Nissim, Ctr. for Research on Computation & Soc'y, Presentation at Harvard University: Differential Privacy in CDFs (Apr. 2016), http://privacytools.seas.harvard.edu/files/dpcdf_user_manual_aug_2016.pdf [https://perma.cc/DZU8-7SSB] (slide deck).

111. *See* WILLIAM H. GREEN, ECONOMETRIC ANALYSIS 13–14, 28–29 (8th ed. 2017).

112. *See id.*

113. *See, e.g.*, Adam Smith, *Privacy-Preserving Statistical Estimation with Optimal Convergence Rates*, 43 PROC. ACM SYMP. ON THEORY COMPUTING 813, 814 (2011).

114. *See* TREVOR HASTIE, ROBERT TIBSHIRANI & JEROME FRIEDMAN, THE ELEMENTS OF STATISTICAL LEARNING: DATA MINING, INFERENCE, & PREDICTION 501 (2d ed. 2001).

115. *See id.* at 502.

116. Many papers describe differentially private clustering algorithms. For a recent example, see Haim Kaplan & Uri Stemmer, *Differentially Private k-Means with Constant Multiplicative Error* 1 (ArXiv, Working Paper No. 1804.08001, 2018), https://arxiv.org/abs/1804.08001 [https://perma.cc/HR35-FHHK].

to be included in future privacy-preserving tool kits for social scientists.

- **Classification:** In machine learning and statistics, classification is the problem of identifying or predicting which of a set of categories a data point belongs in, based on a training set of examples for which category membership is known.[117] Data scientists often utilize data samples that are pre-classified (e.g., by experts or from historical data) to train a classifier, which can later be used for labeling newly acquired data samples.[118] Theoretical work has shown that it is possible to construct differentially private classification algorithms for a large collection of classification tasks.[119]
- **Synthetic data:** Synthetic data are data sets generated from a statistical model estimated using the original data.[120] The records in a synthetic data set have no one-to-one correspondence with the individuals in the original data set, yet the synthetic data can retain many of the statistical properties of the original data. Synthetic data resemble the original sensitive data in format, and, for a large class of analyses, results are similar whether performed on the synthetic or original data.[121] Theoretical work has shown that differentially private synthetic data can be generated for a large variety of tasks.[122] A significant benefit is that, once a differentially private synthetic data set is generated, it can be analyzed any number of times, without any further implications for privacy.[123] As a result, synthetic data can be shared freely

117. *See* JAMES ET AL., *supra* note 102, at 127–29.

118. *See id.*

119. Many papers describe differentially private classification algorithms. For an early example, see Blum et al., *supra* note 46.

120. *See* Jerome P. Reiter, *Satisfying Disclosure Restrictions with Synthetic Data Sets*, 18 J. OFFICIAL STAT. 531, 531 (2002); Jerome P. Reiter & Trivellore E. Raghunathan, *The Multiple Adaptations of Multiple Imputation*, 102 J. AM. STAT. ASS'N 1462, 1466 (2007); Donald B. Rubin, Discussion, *Statistical Disclosure Limitation*, 9 J. OFFICIAL STAT. 461, 464 (1993).

121. *See* Rubin, *supra* note 120, at 463.

122. *See, e.g.*, Avrim Blum, Katrina Ligett & Aaron Roth, *A Learning Theory Approach to Non-Interactive Database Privacy*, 40 PROC. ACM SYMP. ON THEORY COMPUTING 609, 609 (2008).

123. *See* NAT'L ACADS. OF SCIS., ENG'G & MED., INNOVATIONS IN FEDERAL STATISTICS: COMBINING DATA SOURCES WHILE PROTECTING PRIVACY 94 (Robert M. Groves & Brian A. Harris-Kojetin eds., 2017).

or even made public in many cases.[124] For example, statistical agencies can release synthetic microdata as public-use data files in place of raw microdata.[125]

VI. PRACTICAL CONSIDERATIONS WHEN USING DIFFERENTIAL PRIVACY

This Part discusses some of the practical challenges to using differentially private computations such as those outlined in the previous Part. When making a decision regarding whether to implement differential privacy, one must consider the relevant privacy and utility requirements associated with the specific use case in mind. This Article provides many examples illustrating scenarios in which differentially private computations could be used. However, if, for instance, an analysis is being performed at the individual-level—e.g., in order to identify individual patients who would be good candidates for a clinical trial or to identify instances of bank fraud—differential privacy would not apply, as it will disallow learning information specific to an individual.

Additionally, because implementation and use of differential privacy is in its early stages, there is a current lack of easy-to-use general purpose and production-ready tools, though progress is being made on this front, as Part VII discusses below. The literature identifies a number of other practical limitations, emphasizing the need for additional differentially private tools tailored to specific applications such as the data products released by federal statistical agencies; subject matter experts trained in the practice of differential privacy; tools for communicating the features of differential privacy to the general public, users, and other stakeholders; and guidance on setting the privacy loss parameter ε.[126]

This Part focuses on a selection of practical considerations, including (A) challenges due to the degradation of privacy that results from composition, (B) challenges related to the accuracy of differentially private statistics, and (C) challenges related to analyzing and sharing personal data while protecting privacy in accordance with applicable

124. For an example of public use synthetic microdata, see Ashwin Machanavajjhala et al., *Privacy: Theory Meets Practice on the Map*, 24 PROC. IEEE INT'L CONF. ON DATA ENGINEERING 277, 277 (2008).

125. *See* Ron S. Jarmin, Thomas A. Louis & Javier Miranda, *Expanding the Role of Synthetic Data at the U.S. Census Bureau* 3 (Ctr. for Econ. Studies, Research Paper No. CES 14-10, 2014), https://www2.census.gov/ces/wp/2014/CES-WP-14-10.pdf [https://perma.cc/6UXH-TMKM].

126. *See* Simson L. Garfinkel, John M. Abowd & Sarah Powazek, *Issues Encountered Deploying Differential Privacy* (ArXiv, Working Paper No. 1809.02201, 2018), https://arxiv.org/abs/1809.02201 [https://perma.cc/4FL6-JU46].

regulations and policies for privacy protection. It is important to note that the challenges of producing accurate statistics, while protecting privacy and addressing composition, are not unique to differential privacy.[127] It is a fundamental law of information that privacy risk grows with the repeated use of data, and hence this risk applies to any disclosure limitation technique.[128] Traditional SDL techniques—such as suppression, aggregation, and generalization—often reduce accuracy and are vulnerable to loss in privacy due to composition.[129] The impression that these techniques do not suffer accumulated degradation in privacy is merely due to the fact that these techniques have not been analyzed with the high degree of rigor that differential privacy has been.[130] A rigorous analysis of the effect of composition is important for establishing a robust and realistic understanding of how multiple statistical computations affect privacy.[131]

A. The "Privacy Budget"

As Section IV.B explains, one can think of the parameter ε as determining the overall privacy protection provided by a differentially private analysis. Intuitively, ε determines "how much" of an individual's privacy an analysis may utilize, or, alternatively, by how much the risk to an individual's privacy can increase. A smaller value for ε implies better protection (i.e., less risk to privacy).[132] Conversely, a larger value for ε implies worse protection (i.e., higher potential risk to privacy).[133] In particular, $\varepsilon = 0$ implies perfect privacy (i.e., the analysis does not increase any individual's privacy risk at all).[134] Unfortunately, analyses that satisfy differential privacy with $\varepsilon = 0$ must completely ignore their input data and therefore are useless.[135]

Section IV.B also explains that the choice of ε is dependent on various normative and technical considerations, and best practices are

127. *See* Dwork et al., *supra* note 62, at 82.

128. *See id.*

129. *See* Srivatsava Ranjit Ganta, Shiva Prasad Kasiviswanathan & Adam Smith, *Composition Attacks and Auxiliary Information in Data Privacy*, 14 PROC. ACM SIGKDD INT'L CONF. ON KNOWLEDGE, DISCOVERY & DATA MINING 265, 265–66 (2008).

130. For a discussion of privacy and utility with respect to traditional statistical disclosure limitation techniques, see generally Bee-Chung Chen et al., *Privacy-Preserving Data Publishing*, 2 FOUND. & TRENDS IN DATABASES 1 (2009). As shown in Example 5, techniques relying on aggregation do not necessarily compose well. Furthermore, this phenomenon has been demonstrated more generally with respect to a wide range of traditional statistical disclosure limitation techniques. *See generally* Ganta, Kasiviswanathan & Smith, *supra* note 129.

131. *See id.* at 266.

132. *See* Dwork et al., *supra* note 62, at 18.

133. *See id.* at 18.

134. *See id.*

135. *See supra* Part IV.B.

likely to emerge over time as practitioners gain experience from working with real-world implementations of differential privacy. As a starting point, experts have suggested that ε be thought of as a small value ranging from approximately 0.01 to 1.[136] Based on the analysis following Table 1, the Authors of this Article believe that adopting a global value of $\varepsilon = 0.1$, when feasible, provides sufficient protection. In general, setting ε involves making a compromise between privacy protection and accuracy. The consideration of both utility and privacy is challenging in practice and, in some of the early implementations of differential privacy, has led to choosing a higher value for ε.[137] As the accuracy of differentially private analyses improves over time, it is likely that lower values of ε will be chosen.

The privacy loss parameter ε can be thought of as a "privacy budget" to be spent by different analyses of individuals' data. If a single analysis is expected to be performed on a given set of data, then one might allow this analysis to exhaust the entire privacy budget ε. However, a more typical scenario is that several analyses are expected to be run on a dataset, and, therefore, one needs to calculate the total utilization of the privacy budget by these analyses. [138]

Fortunately, as Section IV.E discusses, a number of composition theorems have been developed for differential privacy. In particular, these theorems state that the composition of two differentially private analyses results in a privacy loss that is bounded by the sum of the privacy losses of each of the analyses.[139]

To understand how overall privacy loss is accounted for in this framework, consider the following example.

Example 15

Suppose a data analyst using a differentially private analysis tool is required to do so while maintaining differential privacy with an overall privacy loss parameter $\varepsilon = 0.1$. This requirement for the overall privacy loss parameter may be guided by an interpretation of a regulatory standard, institutional policy, or best practice, among other possibilities. It means that all of the analyst's analyses, taken together, must have a value of ε that is at most 0.1.

136. *See, e.g.*, Dwork, *A Firm Foundation*, *supra* note 46, at 91 ("[W]e tend to think of ε as, say, 0.01, 0.1, or in some cases, ln 2 or ln 3.").

137. *See supra* notes 85–86 and the discussion following Table 1.

138. *See* Heffetz & Ligett, *supra* note 46, at 84 (discussing various examples in which the privacy budget is divided across several analyses).

139. *See* Dwork et al., *supra* note 62, at 28.

Consider how this requirement would play out within the following scenarios:

One-query scenario: The data analyst performs a differentially private analysis with a privacy loss parameter $\varepsilon_1 = 0.1$. In this case, the analyst would not be able to perform a second analysis over the data without risking a breach of the policy limiting the overall privacy loss to $\varepsilon = 0.1$.

Multiple-query scenario: The data analyst first performs a differentially private analysis with $\varepsilon_1 = 0.01$, which falls below the limit of $\varepsilon = 0.1$. This means that the analyst can also apply a second differentially private analysis, say with $\varepsilon_2 = 0.02$. After the second analysis, the overall privacy loss amounts to

$$\varepsilon_1 + \varepsilon_2 = 0.01 + 0.02 = 0.03,$$

which is still less than $\varepsilon = 0.1$, and therefore allows the analyst to perform additional analyses before exhausting the budget.

The multiple-query scenario can be thought of as if the data analyst has a privacy budget of $\varepsilon = 0.1$ that is consumed incrementally as she performs differentially private analyses, until the budget has been exhausted.[140] Performing additional analyses after the overall budget has been exhausted may result in a privacy parameter that is larger (i.e., worse) than ε.[141] Any data use exceeding the privacy budget would result in a privacy risk that is too significant.

Note that, in the sample calculation for the multiple-query example, the accumulated privacy risk was bounded simply by adding the privacy parameters of each analysis. It is in fact possible to obtain better bounds on the accumulation of the privacy loss parameter than suggested by this example.[142] Various tools for calculating the bounds on the accumulated privacy risks in real-world settings using more sophisticated approaches are currently under development.[143]

140. *See* Heffetz & Ligett, *supra* note 46, at 84.

141. *See id.* at 84, 87.

142. A number of papers explore ways to improve these bounds. *See, e.g.*, Amos Beimel, Kobbi Nissim & Eran Omri, *Distributed Private Data Analysis: Simultaneously Solving How and What*, 2008 ADVANCES IN CRYPTOGRAPHY (CRYPTO) 451; Cynthia Dwork, Guy N. Rothblum & Salil Vadhan, *Boosting and Differential Privacy*, 51 IEEE ANN. SYMP. ON FOUND, COMPUTER SCI. 51 (2010); Peter Kairouz, Sewoong Oh & Pramod Viswanath, *The Composition Theorem for Differential Privacy*, 63 IEEE TRANSACTIONS ON INFO. THEORY 4037 (2017); Jack Murtagh & Salil P. Vadhan, *The Complexity of Computing the Optimal Composition of Differential Privacy*, 2016 THEORY OF CRYPTOGRAPHY 157.

143. *See* Gaboardi et al., *supra* note 78, at 7.

B. Accuracy

This Section discusses the relationship between differential privacy and accuracy. The accuracy of an analysis is a measure of how its outcome can deviate from the true quantity or model it attempts to estimate.[144] There is no single measure of accuracy, as measures of deviations differ across applications.[145] Multiple factors have an effect on the accuracy of an estimate, including measurement and sampling errors.[146] The random noise introduced in differentially private computations similarly affects accuracy.[147]

For most statistical analyses, the inaccuracy coming from sampling error decreases as the number of samples grows,[148] and the same is true for the inaccuracy coming from the random noise in most differentially private analyses. In fact, it is often the case that the inaccuracy due to the random noise vanishes more quickly than the sampling error.[149] This means that, in theory, for very large datasets (with records for very many individuals), differential privacy comes essentially "for free."

However, for datasets of the sizes that occur in practice, the amount of noise that is introduced for differentially private analyses can have a noticeable impact on accuracy. For small datasets, for very high levels of privacy protection (i.e., small ε), or for complex analyses, the noise introduced for differential privacy can severely impact utility.[150] In general, almost no utility can be obtained from datasets containing $1/\varepsilon$ or fewer records.[151] As Section VI.A discusses, this is

144. *See* INT'L STATISTICAL INST., THE OXFORD DICTIONARY OF STATISTICAL TERMS 4 (Yadolah Dodge ed., 6th ed. 2006).

145. For example, a researcher interested in estimating the average income of a given population may care about the absolute error of this estimate (i.e., the difference between the real average and the estimate), whereas a researcher interested in the median income may care about the difference between the number of respondents whose income is below the estimate and the number of respondents whose income is above the estimate.

146. Measurement error is the difference between the measured value of a quantity and its true value (e.g., an error in measuring an individual's height or weight), and sampling error is error caused by observing a sample rather than the entire population (e.g., the fraction of people with diabetes in the sample is likely to be different from the fraction with diabetes in the population).

147. *See* Muise & Nissim, *supra* note 110, at 94.

148. *See* JACOB COHEN, STATISTICAL POWER ANALYSIS FOR THE BEHAVIORAL SCIENCES 6 (1977).

149. *See generally* Dwork et al., *supra* note 62; Smith, *supra* note 113; *infra* Appendix A.2.

150. *See* Muise & Nissim, *supra* note 110; Michael Hay et al., *Principled Evaluation of Differentially Private Algorithms Using DPBench*, 2016 PROC. ACM SIGMOD INT'L CONF. ON MGMT. DATA 139, 139, http://dl.acm.org/citation.cfm?id=2882931 [https://perma.cc/6BQD-PQCT].

151. This rule of thumb follows directly from the definition of differential privacy. *See* Dwork et al., *supra* note 62, at 17, 18. Specifically, the parameter ε bounds the distance between the probability distributions resulting from a differentially private computation on two datasets that differ on one entry. Datasets containing only $1/\varepsilon$ entries can differ on at most this number of

exacerbated by the fact that the privacy budget usually needs to be partitioned among many different queries or analyses, and thus the value of ε used for each query needs to be much smaller. Much of the ongoing research on differential privacy is focused on understanding and improving the tradeoff between privacy and utility (i.e., obtaining the maximum possible utility from data while preserving differential privacy).[152]

Procedures for estimating the accuracy of certain types of analyses have been developed.[153] These procedures take as input the number of records, a value for ε, and the ranges of numerical and categorical fields, among other parameters, and produce guaranteed accuracy bounds.[154] Alternatively, a desired accuracy may be given as input instead of ε, and the computation results in a value for ε that would provide this level of accuracy.[155] Figures 4(a)–(d) illustrate an example of a cumulative distribution function and the results of its noisy approximation with different settings of the privacy parameter ε.[156]

entries. Summing the differences over just $1/\varepsilon$ entries reveals that, for any two datasets of this size, the differentially private mechanism produces distributions that are at distance $\varepsilon \cdot \frac{1}{\varepsilon} = 1$ at most. A distance of this size would usually not support any reasonable utility.

152. *See, e.g.*, Dwork, *Differential Privacy*, *supra* note 46, at 6; DWORK & ROTH, *supra* note 25, at 158; Vadhan, *supra* note 46, at 58–59, 77.

153. *See* Mohan et al., *supra* note 78, at 349; Gaboardi et al., *supra* note 78, at 15.

154. *See* Gaboardi et al., *supra* note 78, at 15.

155. *See id.* at 12, 15.

156. Figures 4(a)–(d) are adapted from Muise & Nissim, *supra* note 110, at 113.

Figure 4. Example of the Differentially Private Computation Output

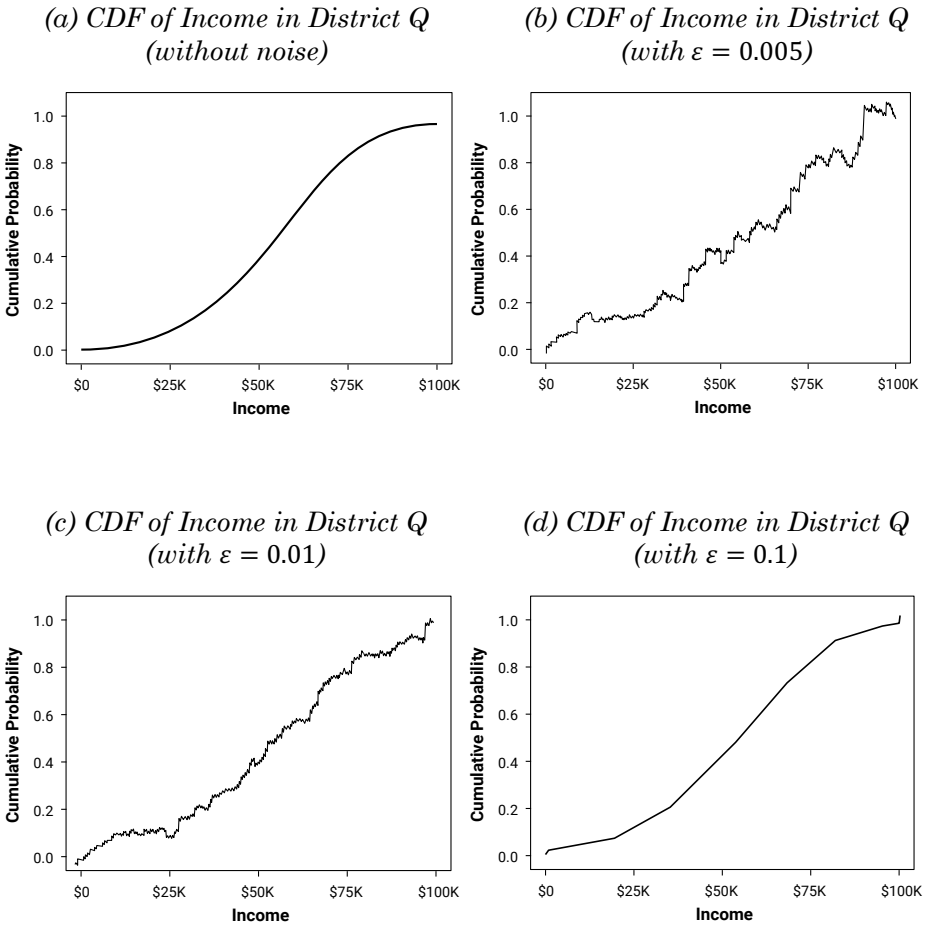


Figure 4 illustrates the outcome of a differentially private computation of the CDF of income in fictional District Q. Graph (a) presents the original CDF (without noise) and the subsequent graphs show the result of applying differentially private computations of the CDF with ε values of (b) 0.005, (c) 0.01, and (d) 0.1. Notice that, as smaller values of ε imply better privacy protection, they also imply less accuracy due to noise addition compared to larger values of ε.

Another concept related to accuracy is truthfulness. This term has appeared regularly, if infrequently, in the statistical disclosure limitation literature since the mid-1970s, though it does not have a

well-recognized formal definition.[157] Roughly speaking, the SDL literature recognizes a privacy-protecting method as truthful if one can determine unambiguously which types of statements, when semantically correct as applied to the protected data (i.e., data transformed by a privacy technique such as k-anonymity), are also semantically correct when applied to the original sample data.[158]

This concept has an intuitive appeal. For data protected via suppressing some of the cells in the database, statements of the form "there are records with characteristics X and Y" are correct in the original data if they are correct in the protected data. For example, one might definitively state, using only the protected data, that "some plumbers earn over $50,000." One cannot make this same statement definitively for data that have been synthetically generated.[159]

One must be careful, however, to identify and communicate the types of true statements a protection method supports. For instance, neither suppression nor synthetic data support truthful nonexistence claims at the microdata level. Even if all Wisconsin residents are included in the data, a statement such as "there are no plumbers in the dataset who earn over $50,000" cannot be made definitively by examining the protected data alone if income or occupation values have been suppressed or synthetically generated. Moreover, protection methods may, in general, preserve truth at the individual record level, but not at the aggregate level (or vice versa).[160] For instance, local

157. *See, e.g.*, Lawrence H. Cox & Gordon Sande, *Techniques for Preserving Statistical Confidentiality*, 42 PROC. INT'L STAT. INST. 6 (1979); Josep Domingo-Ferrer, David Sánchez & Jordi Soria-Comas, *Database Anonymization: Privacy Models, Data Utility, and Microaggregation-Based Inter-Model Connections*, 15 SYNTHESIS LECTURES INFO. SECURITY, PRIVACY & TR. 1, 15 (2016) (distinguishing between "perturbative masking (which distorts the original data and leads to the publication of non-truthful data) and non-perturbative masking (which reduces the amount of information, either by suppressing some of the data or by reducing the level of detail, but preserves truthfulness)"); Benjamin C. M. Fung et al., *Privacy Preserving Data Publication: A Survey of Recent Developments*, 42 ACM COMPUTING SURVS., no. 14, 2010, at 4 (describing, without defining, truthfulness at the record level by explaining that "[i]n some data publishing scenarios, it is important that each published record corresponds to an existing individual in real life. . . . Randomized and synthetic data do not meet this requirement. Although an encrypted record corresponds to a real life patient, the encryption hides the semantics required for acting on the patient represented.").

158. *See* sources cited *supra* note 157. Note that this definition of truthfulness is analogous to the general notion of avoiding false precision and is consistent with recognized principles for reporting statistical results. *See, e.g.*, Tom Lang & Douglas Altman, *Statistical Analyses and Methods in the Published Literature: The SAMPL Guidelines*, 25 Medical Writing 31 (2016).

159. Synthetic data generation, by definition, uses a statistical model built from one set of data to generate new data. This preserves some of the statistical characteristics of the data, but not the original records themselves. *See* Fung et al., *supra* note 157, at 4. As a result, any measurement made on the synthetic dataset is related only probabilistically to measurements made on the original data and is associated with a measure of uncertainty.

160. *See generally* A. F. Karr et al., *A Framework for Evaluating the Utility of Data Altered to Protect Confidentiality*, 60 AM. STATISTICIAN 224 (2006) (discussing various approaches to evaluating the utility of data protected by statistical disclosure limitation techniques).

recoding and suppression, global recoding, and privacy criteria such as k-anonymity that use these operations in their implementation cannot produce reliably truthful statements about most aggregate computations. As an example, statements such as "the median income of a plumber in Wisconsin is $45,000" or "the correlation between income and education in Wisconsin is .50" will not be correct.[161]

Assessing the truthfulness of modern privacy protection methods requires generalizing notions of truthfulness to apply to statements about the population from which the sample is drawn. Scientific research and the field of statistics are primarily concerned with making correct statements about the population.[162] Statistical estimates inherently involve uncertainty and, as mentioned above, there are many individual sources of error that contribute to the total uncertainty in a calculation. These are traditionally grouped by statisticians into the categories of sampling and nonsampling errors.[163] Correct assertions about a statistical statement accurately communicate the uncertainty of the estimated value.[164]

Thus, a statement is statistically truthful of protected data if it accurately communicates the uncertainty—inclusive of sampling and nonsampling errors—of the estimated population value. Methods such as local suppression and global recoding are not always capable of producing statistically truthful statements.[165] Fortunately, privacy

---

161. Correctly calculating and truthfully reporting the uncertainty induced by suppression would require revealing the full details of the suppression algorithm and its parameterization. Revealing these details allows information to be inferred about individuals. Traditional SDL techniques require that the mechanism itself be kept secret in order to protect against this type of attack.

162. In general terms, the goal of statistics is to make reliable inferences about a population or distribution based on characteristics calculated from a sample of data drawn from that population. For a mathematically detailed definition, see Allan Birnbaum, *On the Foundations of Statistical Inference*, 57 J. AM. STAT. ASS'N 269, 273 (1962). In similarly general terms, the goal of science is to yield reliable generalized knowledge about the world, such as knowledge about populations, general predictions, or natural laws. A widely recognized example capturing this distinction is the regulatory definition of scientific research found in the Federal Policy for the Protection of Human Subjects. *See* 45 C.F.R. § 46.102(l) (2018) ("Research means a systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge.").

163. *See Error Measurement,* BUREAU OF LAB. STAT., https://www.bls.gov/opub/hom/topic/error-measurements.htm [https://perma.cc/66U6-HJFA] (last visited Sept. 13, 2018).

164. *See* MICAH ALTMAN, JEFF GILL & MICHAEL P. MCDONALD, NUMERICAL ISSUES IN STATISTICAL COMPUTING FOR THE SOCIAL SCIENTIST 260–61 (2004).

165. *See* LEON WILLENBORG & TON DE WAAL, ELEMENTS OF STATISTICAL DISCLOSURE CONTROL 28 (2001) (discussing how SDL techniques may introduce bias). For instance, Willenborg and de Waal note specifically that suppression of local values (i.e., cells, when used in the context of microdata) induces missing-data bias. Generalization takes many forms, and these forms are associated with different sources of statistical bias. For example, range generalization (e.g., top-coding) involves collapsing the observed distribution of values, which statisticians recognize as yielding truncation bias, whereas global recoding to suppress an entire measure may induce

protecting methods such as synthetic data generation, record swapping, and differential privacy are capable of producing statements about statistical estimates that are truthful.[166] For example, all of these methods could produce truthful statements such as "with a confidence level of 99%, the median income of a plumber is $\$45{,}000 \pm \$2{,}000$."[167] When produced by a truthful method, this statement correctly communicates the uncertainty of the statement, and would, roughly speaking,[168] turn out to be true of the population in 99 out of 100 independent trials.

Generally, differentially private methods introduce uncertainty. However, it is a property of differential privacy that the method itself does not need to be kept secret. This means the amount of noise added to the computation can be taken into account in the measure of accuracy and, therefore, lead to correct statements about the population of interest. This can be contrasted with many traditional SDL techniques, which only report sampling error and keep the information needed to estimate the "privacy error" secret. Any privacy-preserving method, if misused or misinterpreted, can produce incorrect statements. Additionally, the truthfulness of some methods, such as suppression and synthetic data generation, is inherently limited to particular levels of computations (e.g., to existence statements on microdata, or statements about selected aggregate statistical properties, respectively). Differential privacy may be used truthfully for a broader set of computations, so long as the uncertainty of each calculation is estimated and reported.

C. Complying with Legal Requirements for Privacy Protection

Statistical agencies, companies, researchers, and others who collect, process, analyze, store, or share data about individuals must take steps to protect the privacy of the data subjects in accordance with various laws, institutional policies, contracts, ethical codes, and best

---

missing-variable bias in a subsequently estimated model. *See generally* JACK JOHNSTON & JOHN DINARDO, ECONOMETRIC METHODS (4th ed. 1996) (discussing these types of biases).

166. Each of these methods can be applied in such a way that correctly calibrated measures of uncertainty accompany computed statistics. For a detailed treatment of using differential privacy to carefully calibrate the uncertainty in statistical estimates, see Cynthia Dwork et al., *The Reusable Holdout: Preserving Validity in Adaptive Data Analysis*, 349 SCI. 636 (2015).

167. From this statement, we can derive other conclusions, such as that, with 99% confidence, at least half of all plumbers earn over $43,000 annually. And if existence statements such as these are the main concern, one could use other differentially private algorithms to support making similar statements with near certainty—not merely 99% confidence.

168. For a precise treatment of frequentist statistical confidence intervals, see D.R. COX & D.V. HINKLEY, THEORETICAL STATISTICS 48–49, 208–09 (1974).

practices.[169] In some settings, tools that satisfy differential privacy can be used to analyze and share data, while both complying with legal obligations and providing strong mathematical guarantees of privacy protection for the individuals in the data.[170]

Privacy regulations and related guidance do not directly answer the question of whether the use of differentially private tools is sufficient to satisfy existing regulatory requirements for protecting privacy when sharing statistics based on personal data.[171] This issue is complex because privacy laws are often context dependent, and there are significant gaps between differential privacy and the concepts underlying regulatory approaches to privacy protection.[172] Different regulatory requirements are applicable depending on the jurisdiction, sector, actors, and types of information involved.[173] As a result, datasets held by an organization may be subject to different requirements. In some cases, similar or even identical datasets may be subject to different requirements when held by different organizations.[174] In addition, many legal standards for privacy protection are, to a large extent, open to interpretation and therefore require a case-specific legal analysis by an attorney.[175]

Other challenges arise as a result of differences between the concepts appearing in privacy regulations and those underlying differential privacy. For instance, many laws focus on the presence of "personally identifiable information" or the ability to "identify" an individual's personal information in a release of records.[176] Such concepts do not have precise definitions,[177] and their meaning in the context of differential privacy applications is especially unclear.[178] In addition, many privacy regulations emphasize particular requirements for protecting privacy when disclosing individual-level data, such as removing personally identifiable information, which are arguably difficult to interpret and apply when releasing aggregate statistics.[179] While in some cases it may be clear whether a regulatory standard has been met by the use of differential privacy, in other cases—particularly

169. *See supra* Section I.A (discussing legal and ethical frameworks for data privacy).

170. *See* Kobbi Nissim et al., *Bridging the Gap Between Computer Science and Legal Approaches to Privacy*, 31 HARV. J.L. & TECH. 687, 697 (2018).

171. *See id.* at 733.

172. *See id.* at 730, 735.

173. *See id.* at 691; Schwartz & Solove, *supra* note 9, at 1847.

174. *See* Micah Altman et al., *Towards a Modern Approach to Privacy-Aware Government Data Releases*, 30 BERKELEY TECH. L.J. 1967, 2009 (2015).

175. *See id.* at 1972.

176. *See* Schwartz & Solove, *supra* note 9, at 1816.

177. *See id.*

178. *See* Nissim et al., *supra* note 170, at 691, 730–31.

179. *See id.* at 720.

along the boundaries of a standard—there may be considerable uncertainty.[180] Regulatory requirements relevant to issues of privacy in computation rely on an understanding of a range of different concepts, such as personally identifiable information, de-identification, linkage, inference, risk, consent, opt out, and purpose and access restrictions. The following discussion explains how the definition of differential privacy can be interpreted to address each of these concepts while accommodating differences in how these concepts are defined across various legal and institutional contexts.

Personally identifiable information (PII) and de-identification are central concepts in information privacy law.[181] Regulatory protections typically extend only to personally identifiable information; information not considered personally identifiable is not protected.[182] Although definitions of personally identifiable information vary, they are generally understood to refer to the presence of pieces of information that are linkable to the identity of an individual or to an individual's personal attributes.[183] PII is also related to the concept of de-identification, which refers to a collection of techniques devised for transforming identifiable information into non-identifiable information while also preserving some utility of the data. In principle, it is intended that de-identification, if performed successfully, can be used as a tool for removing PII, or transforming PII into non-PII.[184]

When differential privacy is used, it can be understood as ensuring that using an individual's data will not reveal essentially any personally identifiable information specific to her.[185] Here, the use of the term "specific" refers to information that is unique to the individual

180. *See id.* at 710.

181. *See* Schwartz & Solove, *supra* note 9, at 1819.

182. *See id.* at 1816.

183. For a survey of various definitions of *personally identifiable information*, see *id.* at 1829–36. The Government Accountability Office also provides a general definition of personally identifiable information. *See* U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-08-536, ALTERNATIVES EXIST FOR ENHANCING PROTECTION OF PERSONALLY IDENTIFIABLE INFORMATION (2008) ("For purposes of this report, the terms *personal information* and *personally identifiable information* are used interchangeably to refer to any information about an individual maintained by an agency, including (1) any information that can be used to distinguish or trace an individual's identity, such as name, Social Security number, date and place of birth, mother's maiden name, or biometric records; and (2) any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information."), https://www.gao.gov/new.items/d08536.pdf [https://perma.cc/9DTU-H7S6].

184. *See, e.g.*, 34 C.F.R. § 99.31(b)(1) (2018) (provision for "[d]e-identified records and information," which permits the release of education records "after the removal of all personally identifiable information provided that the educational agency or institution or other party has made a reasonable determination that a student's identity is not personally identifiable, whether through single or multiple releases, and taking into account other reasonably available information").

185. Note that the reference to "using an individual's data" in this statement means the inclusion of an individual's data in an analysis.

and cannot be inferred unless the individual's information is used in the analysis.

Linkage is a mode of privacy loss recognized, implicitly or explicitly, by a number of privacy regulations.[186] As illustrated in Example 1, linkage typically refers to the matching of information in a database to a specific individual, often by leveraging information from external sources.[187] Linkage is also closely related to the concept of identifying an individual in a data release, as identifying an individual is often accomplished via a successful linkage.[188] Linkage has a concrete meaning when data are published as a collection of individual-level records, often referred to as microdata.[189] However, what is considered a successful linkage when a publication is made in other formats, such as statistical models or synthetic data, has not been defined and is open to interpretation.

Despite this ambiguity, it can be argued that differential privacy addresses record linkage in the following sense. Differentially private statistics provably hide the influence of every individual, and even small groups of individuals.[190] Although linkage has not been precisely defined, linkage attacks seem to inherently result in revealing that specific individuals participated in an analysis. Because differential privacy protects against learning whether or not an individual participated in an analysis, it can therefore be understood to protect against linkage. Furthermore, differential privacy provides a robust guarantee of privacy protection that is independent of the auxiliary information available to an attacker.[191] Indeed, under differential privacy, even an attacker utilizing arbitrary auxiliary information cannot learn much more about an individual in a database than she could if that individual's information were not in the database at all.[192]

---

186. For example, by defining personally identifiable information in terms of information "linked or linkable to a specific student," FERPA appears to emphasize the risk of a successful record linkage attack. *See* 34 C.F.R. § 99.3 (2018). The Department of Health & Human Services in guidance on de-identifying data in accordance with the HIPAA Privacy Rule includes an extended discussion of examples of record linkage attacks and de-identification strategies for mitigating them. *See* DEP'T OF HEALTH & HUMAN SERVS., *supra* note 10, at 15–17. Guidance on complying with European data protection law refers to linkability, "which is the ability to link, at least, two records concerning the same data subject or a group of data subjects (either in the same database or in two different databases)," as one of three risks essential to anonymization. *Article 29 Data Protection Working Party*, *Opinion 05/2014 on Anonymisation Techniques*, at 11 (Apr. 10, 2014) [hereinafter *Article 29 Data Protection Working Party*].

187. *See* DWORK & ROTH, *supra* note 25, at 6–7; Fed. Comm. on Statistical Methodology, *supra* note 19, at 83.

188. *See* sources cited *infra* note 186.

189. *See* Fed. Comm. on Statistical Methodology, *supra* note 19, at 4.

190. *See* Dwork et al., *supra* note 62, at 17, 29.

191. *See* Ganta, Kasiviswanathan & Smith, *supra* note 129, at 265.

192. *See id.* at 271.

Inference is another mode of privacy loss that is implicitly or explicitly referenced by some privacy regulations and related guidance. For example, some laws protect information that enables the identity of an individual to be "reasonably inferred,"[193] and others protect information that enables one to determine an attribute about an individual with "reasonable certainty."[194] When discussing inference as a mode of privacy loss, it is important to distinguish between two types—inferences about individuals and inferences about large groups of individuals. Although privacy regulations and related guidance generally do not draw a clear distinction between these two types of inference,[195] the distinction is key to understanding which privacy safeguards would be appropriate in a given setting.

Differential privacy can be understood as essentially protecting an individual from inferences about attributes that are specific to her—that is, information that is unique to the individual and cannot be inferred unless the individual's information is used in the analysis. Interventions other than differential privacy may be necessary in contexts in which inferences about large groups of individuals, such as uses of data that result in discriminatory outcomes by race or sex, are a concern.[196]

Risk is another concept that appears in various ways throughout regulatory standards for privacy protection and related guidance. For example, some regulatory standards include a threshold level of risk that an individual's information may be identified in a data release.[197] Similarly, some regulations also acknowledge, implicitly or explicitly, that any disclosure of information carries privacy risks, and therefore the goal is to minimize, rather than eliminate, such risks.[198]

---

193. *See, e.g.*, E-Government Act of 2002, Pub. L. 107-347, 116 Stat. 2899, § 208 (2002) (codified as amended at 44 U.S.C. § 3501 (2012)) ("[T]he term 'identifiable form' means any representation of information that permits the identity of an individual to whom the information applies to be reasonably inferred by either direct or indirect means.").

194. *See, e.g.*, 34 C.F.R. § 99.3 (2018) (defining "personally identifiable information," in part, in terms of information that would allow one to identify a student "with reasonable certainty").

195. *See, e.g.*, *Article 29 Data Protection Working Party*, *supra* note 186, at 12 (defining inference broadly as "the possibility to deduce, with significant probability, the value of an attribute from the values of a set of other attributes").

196. *See* Micah Altman et al., *Practical Approaches to Big Data Privacy Over Time*, 8 INT'L DATA PRIVACY L. 29, 43 (2018); Micah Altman, Alexandra Wood & Effy Vayena, *A Harm-Reduction Framework for Algorithmic Fairness*, 16 IEEE SECURITY & PRIVACY 34 (2018).

197. The HIPAA Privacy Rule requires covered entities to use de-identification techniques prior to releasing data in order to create a dataset with only a "very small" risk of identification. 45 C.F.R. § 164.514(b)(1) (2018).

198. Guidance on complying with the Confidential Information Protection and Statistical Efficiency Act (CIPSEA) requires agencies to "[c]ollect and handle confidential information to minimize risk of disclosure." *See* Implementation Guidance for Title V of the E-Government Act, 72 Fed. Reg. 33,362–33,363 (June 15, 2007). Guidance from the Department of Health & Human

Differential privacy can readily be understood in terms of risk.[199] Specifically, differential privacy enables a formal quantification of risk.[200] It guarantees that the risk to an individual is essentially the same with or without her participation in the dataset,[201] and this is likely true for most notions of risk adopted by regulatory standards or institutional policies. In this sense, differential privacy can be interpreted as essentially guaranteeing that the risk to an individual is minimal or very small. Moreover, the privacy loss parameter ε can be tuned according to different requirements for minimizing risk.[202]

Consent and opt out are concepts underlying common provisions set forth in information privacy laws.[203] Consent and opt-out provisions enable individuals to choose to allow, or not to allow, their information to be used by or redisclosed to a third party.[204] Such provisions are premised on the assumption that providing individuals with an opportunity to opt in or out gives them control over the use of their personal information and effectively protects their privacy.[205] However, this assumption warrants a closer look. Providing consent or opt-out mechanisms as a means of providing individuals with greater control over their information is an incomplete solution as long as individuals are not fully informed about the consequences of uses or disclosures of their information.[206] In addition, allowing individuals the choice to opt in or out can create new privacy concerns. For example, an individual's decision to opt out may—often unintentionally—be reflected in a data release or analysis and invite scrutiny into whether the choice to opt out was motivated by the need to hide compromising information.[207]

The differential privacy guarantee can arguably be interpreted as providing stronger privacy protection than a consent or opt-out mechanism. This is because differential privacy can be understood as

---

Services recognizes that de-identification methods "even when properly applied, yield de-identified data that retains some risk of identification. Although the risk is very small, it is not zero, and there is a possibility that de-identified data could be linked back to the identity of the patient to which it corresponds." DEP'T OF HEALTH & HUMAN SERVS., *supra* note 10, at 6.

199. *See supra* Section IV.C.

200. *See id.*

201. *See id.*

202. *See supra* Section IV.B.

203. *See generally* Daniel J. Solove, *Introduction: Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1880, 1884, 1901 (2013).

204. *See, e.g.*, 34 C.F.R. § 99.37 (2018) (including a provision requiring educational agencies and institutions to offer students an opportunity to opt out of the disclosure of their personal information in school directories).

205. *See* Solove, *supra* note 203, at 1880.

206. *See id.* at 1885.

207. *See, e.g.*, Kim Zetter, *The NSA Is Targeting Users of Privacy Services, Leaked Code Shows*, WIRED (July 3, 2014, 5:45 PM), https://www.wired.com/2014/07/nsa-targets-users-of-privacy-services/ [https://perma.cc/2KVL-LKS4] (revealing that the National Security Agency's surveillance efforts specially target users of privacy services).

automatically providing all individuals in the data with essentially the same protection that opting out is intended to provide.[208] Moreover, differential privacy provides all individuals with this privacy guarantee.[209] Therefore, differential privacy can be understood to prevent the possibility that individuals who choose to opt out would, by doing so, inadvertently reveal a sensitive attribute about themselves or attract attention as individuals who are potentially hiding sensitive facts about themselves.

Purpose and access provisions often appear in privacy regulations as restrictions on the use or disclosure of personal information to specific parties or for specific purposes. Legal requirements reflecting purpose and access restrictions can be divided into two categories. The first category includes restrictions, such as those governing confidentiality for statistical agencies,[210] prohibiting the use of identifiable information except for statistical purposes. The second category broadly encompasses other types of purpose and access provisions, such as those permitting the use of identifiable information for legitimate educational purposes.[211]

Restrictions limiting use to statistical purposes, including statistical purposes involving population-level rather than individual-level analyses or statistical computations, are in many cases consistent with the use of differential privacy. This is because, as Part IV explains, differential privacy protects information specific to an individual while allowing population-level analyses to be performed. Therefore, tools that satisfy differential privacy may be understood to restrict uses to only those that are for statistical purposes, such as the definition of statistical purposes found in the Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA).[212] However, other use and access restrictions, such as provisions limiting use to legitimate educational purposes, are orthogonal to differential privacy and require alternative privacy safeguards.[213]

---

208. *See supra* Part IV.

209. *See id.*

210. *See, e.g.*, Confidential Information Protection and Statistical Efficiency Act of 2002, Pub. L. No. 107-347, 116 Stat. 2899, 2963, 2966 (2002) (codified as amended at 44 U.S.C. § 3501 (2012)) (prohibiting the use of protected information "for any use other than an exclusively statistical purpose," where *statistical purpose* "means the description, estimation, or analysis of the characteristics of groups, without identifying the individuals or organizations that comprise such groups").

211. For example, FERPA generally prohibits the disclosure of personally identifiable information from education records, with limited exceptions such as disclosures to school officials with a legitimate educational interest in the information, 34 C.F.R. § 99.31(a)(1) (2018), or to organizations conducting studies for, or on behalf of, schools, school districts, or postsecondary institutions, § 99.31(a)(6).

212. *See supra* note 210.

213. *See* Altman et al., *supra* note 196, at 47.

The foregoing interpretations of the differential privacy guarantee can be used to demonstrate that, in many cases, a differentially private mechanism would prevent the types of disclosures of personal information that privacy regulations have been designed to address. Moreover, in many cases, differentially private tools provide privacy protection that is more robust than that provided by techniques commonly used to satisfy regulatory requirements for privacy protection. However, further research to develop methods for proving that differential privacy satisfies legal requirements and setting the privacy loss parameter ε based on such requirements is needed.[214] In practice, data providers should consult with legal counsel when considering whether differential privacy tools—potentially in combination with other tools for protecting privacy and security—are appropriate within their specific institutional settings.[215]

VII. TOOLS FOR DIFFERENTIALLY PRIVATE ANALYSIS

At the time of this writing, differential privacy is transitioning from a purely theoretical mathematical concept to one that underlies software tools for practical use by analysts of privacy-sensitive data. The first real-world implementations of differential privacy have been deployed by companies such as Google,[216] Apple,[217] and Uber,[218] and government agencies such as the US Census Bureau.[219] Researchers in industry and academia are currently building and testing additional tools for differentially private statistical analysis. This Part briefly reviews some of these newly emerging tools, with a particular focus on the tools that inspired the drafting of this primer.

---

214. For an extended discussion of the gaps between legal and computer science definitions of privacy and a demonstration that differential privacy can be used to satisfy an institution's obligations under FERPA, see Nissim et al., *supra* note 170.

215. For a framework for selecting among differential privacy and other suitable privacy and security controls, see Altman et al., *supra* note 196, at 29; Altman et al., *supra* note 174, at 2022.

216. *See* Úlfar Erlingsson, Vasyl Pihur & Aleksandra Korolova, *RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response*, 2014 PROC. ACM CONF. ON COMPUTER & COMM. SECURITY 1054, 1055 (2014) [hereinafter Erlingsson et al., *RAPPOR*]; Úlfar Erlingsson, *Learning Statistics with Privacy, Aided by the Flip of a Coin*, GOOGLE AI BLOG (Oct. 30, 2014), http://googleresearch.blogspot.com/2014/10/learning-statistics-with-privacy-aided.html [https://perma.cc/Q873-TZZS].

217. Andy Greenberg, *Apple's 'Differential Privacy' Is About Collecting Your Data—But Not Your Data*, WIRED (June 13, 2016, 7:02 PM), http://www.wired.com/2016/06/apples-differential-privacy-collecting-data/ [https://perma.cc/5A47-GP96].

218. *See* Noah Johnson, Joseph P. Near & Dawn Song, *Towards Practical Differential Privacy for SQL Queries*, 11 PROC. VLDB ENDOWMENT 526, 526 (2018).

219. *See OnTheMap Application for the Longitudinal Employer-Household Dynamics Program*, US CENSUS BUREAU, http://onthemap.ces.census.gov [https://perma.cc/WNX3-CQFB] (last visited Sept. 25, 2018).

A. Government and Commercial Applications of Differential Privacy

Since 2006, the US Census Bureau has published an online interface enabling the exploration of the commuting patterns of workers across the United States, based on confidential data collected by the Bureau through the Longitudinal Employer-Household Dynamics program.[220] Through this interface, members of the public can interact with synthetic datasets generated from confidential survey records.[221] Beginning in 2008, the computations used to synthesize the data accessed through the interface have provided formal privacy guarantees that satisfy a variant of differential privacy.[222] In 2017, the Census Bureau announced that it was prototyping a system that would protect the full set of publication products from the 2020 decennial Census using differential privacy.[223]

Google, Apple, and Uber have also experimented with differentially private implementations.[224] For instance, Google developed the RAPPOR system, which applies differentially private computations in order to gather aggregate statistics from consumers who use the Chrome web browser.[225] This tool allows analysts at Google to monitor the wide-scale effects of malicious software on the browser settings of Chrome users, while providing strong privacy guarantees to individuals.[226] The current differentially private implementations by the Census Bureau and Uber rely on a curator model—the model serving as the focus of most of this Article—in which a database administrator has access to and uses private data to generate differentially private data summaries.[227] In contrast, the current implementations by Google's RAPPOR and in Apple's macOS 10.12 and iOS 10 rely on a local model of privacy, which does not require individuals to share their private data with a trusted third party; but

220. *See id.*

221. *See OnTheMap Help and Documentation*, US CENSUS BUREAU, https://lehd.ces.census.gov/applications/help/onthemap.html#!faqs [https://perma.cc/P7PU-4CL2] (last visited Oct. 4, 2018).

222. *See* Machanavajjhala et al., *supra* note 124, at 277.

223. *See generally* Garfinkel, Abowd & Powazek, *supra* note 126.

224. *See* Erlingsson et al., *RAPPOR*, *supra* note 216; Greenberg, *supra* note 217; Johnson, Near & Song, *supra* note 218, at 526.

225. *See* Erlingsson et al., *RAPPOR*, *supra* note 216.

226. *Id.* Other examples for using differential privacy (for which, to the best of the Authors' knowledge, no technical reports have been published) include Google's use of differential privacy in analyzing urban mobility and Apple's use of differential privacy in iOS 10. *See* Andrew Eland, *Tackling Urban Mobility with Technology*, GOOGLE EUR. BLOG (Nov. 18, 2015), http://googlepolicyeurope.blogspot.com/2015/11/tackling-urban-mobility-with-technology.html; Greenberg, *supra* note 217.

227. *See* Garfinkel, Abowd & Powazek, *supra* note 126; Johnson, Near & Song, *supra* note 218.

rather, answer questions about their own data in a differentially private manner.[228] Each of these differentially private answers is not useful on its own, but many of them can be aggregated to perform useful statistical analysis.

B. Research and Development Towards Differentially Private Tools

Several experimental systems from academia and industry enable data analysts to construct privacy-preserving analyses without requiring an understanding of the subtle technicalities of differential privacy. Systems such as Privacy Integrated Queries (PINQ),[229] Airavat,[230] GUPT,[231] Fuzz,[232] DFuzz,[233] and Ektelo[234] aim to provide user-friendly tools for writing programs that are guaranteed to be differentially private, through the use of differentially private building blocks[235] or general frameworks such as "partition-and-aggregate" or "subsample-and-aggregate"[236] for transforming non-private programs into differentially private ones.[237] These systems rely on a common approach: they keep the data safely stored and allow users to access them only via a programming interface which guarantees differential privacy.[238] They also afford generality, enabling one to design many types of differentially private programs that are suitable for a wide range of purposes.[239] However, it can be challenging for a lay user with limited expertise in programming to make effective use of these systems.[240]

The Authors of this Article are collaborators on the Harvard Privacy Tools Project, which develops tools to help social scientists collect, analyze, and share data while providing privacy protection for

228. *See* Erlingsson et al., *RAPPOR*, *supra* note 216; Greenberg, *supra* note 217.

229. Frank McSherry, *Privacy Integrated Queries: An Extensible Platform for Privacy-Preserving Data Analysis*, 2009 PROC. ACM SIGMOD INT'L CONF. ON MGMT. DATA 19, 19–20.

230. Indrajit Roy et al., *Airavat: Security and Privacy for MapReduce*, USENIX (2010), http://www.usenix.org/events/nsdi10/tech/full_papers/roy.pdf [https://perma.cc/N6FF-8SSB].

231. Mohan et al., *supra* note 78, at 349–50.

232. Jason Reed & Benjamin C. Pierce, *Distance Makes the Types Grow Stronger: A Calculus for Differential Privacy*, 15 ACM SIGPLAN INT'L CONF. ON FUNCTIONAL PROGRAMMING 157 (2010).

233. Marco Gaboardi et al., *Linear Dependent Types for Differential Privacy*, 40 PROC. ANN. ACM SIGPLAN-SIGACT SYMP ON PRINCIPLES PROGRAMMING LANGUAGES 357 (2013).

234. Dan Zhang et al., *EKTELO: A Framework for Defining Differentially-Private Computations*, 2018 PROC, INT'L CONF. ON MGMT. DATA 115.

235. *See* McSherry, *supra* note 229, at 91; Gaboardi et al., *supra* note 78, at 6.

236. *See* Kobbi Nissim, Sofya Raskhodnikova & Adam Smith, *Smooth Sensitivity and Sampling in Private Data Analysis*, 39 PROC. ACM SYMP. ON THEORY COMPUTING 75 (2007).

237. *See* Mohan et al., *supra* note 78, at 354; Roy et al., *supra* note 230.

238. *See* Gaboardi et al., *supra* note 78, at 21.

239. *See id.* at 2, 6.

240. *See id.* at 6.

individual research subjects.[241] To this end, the project seeks to incorporate definitions and algorithmic tools from differential privacy into a private data-sharing interface (PSI) which facilitates data exploration and analysis using differential privacy.[242] PSI is intended to be integrated into research data repositories, such as Dataverse.[243] It will provide researchers depositing datasets into a repository with guidance on how to partition a limited privacy budget among the many statistics to be produced or analyses to be run.[244] It will also provide researchers seeking to explore a dataset available on the repository with guidance on how to interpret the noisy results produced by a differentially private algorithm.[245] Through the differentially private access enabled by PSI, researchers will be able to perform rough preliminary analyses of privacy-sensitive datasets that currently cannot be safely shared.[246] Such access will help researchers determine whether it is worth the effort to apply for full access to the raw data.[247]

C. Tools for Specific Data Releases or Specific Algorithms

There have been a number of successful applications of differential privacy with respect to specific types of data—including data from genome-wide association studies,[248] location history data,[249] data on commuter patterns,[250] mobility data,[251] client-side software data,[252] and data on usage patterns for phone technology.[253] For differentially private releases of each of these types of data, experts in differential privacy have taken care to choose algorithms and allocate privacy budgets with the aim of maximizing utility with respect to the particular data set.[254] Therefore, each of these tools is specific to the type of data it is designed to handle, and such tools cannot be applied in contexts in which the collection of data sources and the structure of the datasets are too heterogeneous to be compatible with such

241. *Harvard University Privacy Tools Project*, HARV. U., https://privacytools.seas.harvard.edu/ [https://perma.cc/ABN6-WVE3] (last visited Oct. 1, 2018).
242. *See* Gaboardi et al., *supra* note 78, at 2.
243. *See id.*
244. *See id.*
245. *See id.* at 15, 19.
246. *See id.* at 2, 7.
247. *See id.* at 7.
248. *See* Xiaoqian Jiang et al., *A Community Assessment of Privacy Preserving Techniques for Human Genomes*, 14 BMC MED. INFORMATICS & DECISION MAKING 1, 1–2 (2014).
249. *See* Eland, *supra* note 226.
250. *See* Machanavajjhala et al., *supra* note 124, at 277.
251. *See* Darakhshan J. Mir et al., *DP-WHERE: Differentially Private Modeling of Human Mobility*, 2013 IEEE INT'L CONF. ON BIG DATA 580, 580–82.
252. *See* Erlingsson et al., *RAPPOR*, *supra* note 216, at 1054.
253. *See* Greenberg, *supra* note 217.
254. *See* Gaboardi et al., *supra* note 78, at 6.

optimizations.[255] Thus, there remains a need for more general-purpose tools such as those described in the previous Section. Beyond these examples, a wide literature on the design of differentially private algorithms describes approaches to performing specific data analysis tasks, including work comparing and optimizing such algorithms across a wide range of datasets. For example, the recent development of DPBench,[256] a framework for standardized evaluation of the accuracy of privacy algorithms, provides a way to compare different algorithms and ways of optimizing them.[257]

VIII. SUMMARY

As the previous Part illustrates, differential privacy is in initial stages of implementation in limited academic, commercial, and government settings, and research is ongoing to develop tools that can be deployed in new applications. As differential privacy is increasingly applied in practice, interest in the topic is growing among legal scholars, policymakers, and other practitioners. This Article provides an introduction to the key features of differential privacy, using illustrations that are intuitive and accessible to these audiences.

Differential privacy provides a formal, quantifiable measure of privacy. It is established by a rich and rapidly evolving theory that enables one to reason with mathematical rigor about privacy risk. Quantification of privacy is achieved by the privacy loss parameter ε, which controls, simultaneously for every individual contributing to the analysis, the deviation between one's opt-out scenario and the actual execution of the differentially private analysis.

This deviation can grow as an individual participates in additional analyses, but the overall deviation can be bounded as a function of ε and the number of analyses performed. This amenability to composition—or the ability to provide provable privacy guarantees with respect to the cumulative risk from successive data releases—is a unique feature of differential privacy.[258] While it is not the only framework that quantifies a notion of risk for a single analysis, it is currently the only framework with quantifiable guarantees on the risk resulting from a composition of several analyses.

---

255. *Id.*

256. *See* Michael Hay et al., *Principled Evaluation of Differentially Private Algorithms Using DPBench*, 2016 PROC. ACM SIGMOD INT'L CONF. ON MGMT. DATA 139, 139, http://dl.acm.org/citation.cfm?id=2882931 [https://perma.cc/6BQD-PQCT].

257. *Id.*; *see also* DPCOMP, https://www.dpcomp.org [https://perma.cc/72CL-86ZN] (last visited Sept. 25, 2018).

258. *See* Ganta, Kasiviswanathan & Smith, *supra* note 129, at 265.

The parameter ε can be interpreted as bounding the excess risk to an individual resulting from her data being used in an analysis (compared to her risk when her data are not being used). Indirectly, the parameter ε also controls the accuracy to which a differentially private computation can be performed. For example, researchers making privacy-sensitive data available through a differentially private tool may, through the interface of the tool, choose to produce a variety of differentially private summary statistics while maintaining a desired level of privacy (quantified by an accumulated privacy loss parameter), and then compute summary statistics with formal privacy guarantees.

Systems that adhere to strong formal definitions like differential privacy provide protection that is robust to a wide range of potential privacy attacks, including attacks that are unknown at the time of deployment.[259] An analyst designing a differentially private data release need not anticipate particular types of privacy attacks, such as the likelihood that one could link particular fields with other data sources that may be available. Differential privacy *automatically* provides a robust guarantee of privacy protection that is independent of the methods and resources used by a potential attacker.

Differentially private tools also have the benefit of transparency, as it is not necessary to maintain secrecy around a differentially private computation or its parameters. This feature distinguishes differentially private tools from traditional de-identification techniques which often require concealment of the extent to which the data have been transformed, thereby leaving data users with uncertainty regarding the accuracy of analyses on the data.

Differentially private tools can be used to provide broad, public access to data or data summaries in a privacy-preserving way. Differential privacy can help enable researchers, policymakers, and businesses to analyze and share sensitive data that cannot otherwise be shared due to privacy concerns. Further, it ensures that they can do so with a guarantee of privacy protection that substantially increases their ability to protect the individuals in the data. This, in turn, can further the progress of scientific discovery and innovation.

APPENDIX A. ADVANCED TOPICS

This Article concludes with some advanced topics for readers interested in exploring differential privacy further. This Appendix explores how differentially private analyses are constructed, explains

259. Here, the term "privacy attacks" refers to attempts to learn private information specific to individuals from a data release.

how the noise introduced by differential privacy compares to statistical sampling error, and discusses the protection differential privacy can provide for small groups of individuals.

A.1. How Are Differentially Private Analyses Constructed?

As indicated in Part IV, the construction of differentially private analyses relies on the careful introduction of uncertainty in the form of random noise. This Section provides a simple example illustrating how a carefully calibrated amount of random noise can be added to the outcome of an analysis in order to provide privacy protection.

Example 16

Consider computing an estimate of the number of HIV-positive individuals in a sample, where the sample contains $n = 10{,}000$ individuals of whom $m = 38$ are HIV-positive. In a differentially private version of the computation, random noise Y is introduced into the count so as to hide the contribution of a single individual. That is, the result of the computation would be $m' = m + Y = 38 + Y$ instead of $m = 38$.

The magnitude of the random noise Y affects both the level of privacy protection provided and the accuracy of the count.[260] Generally, greater uncertainty requires a larger noise magnitude and therefore results in worse accuracy—and vice versa. In designing a release mechanism like the one described in Example 16, the magnitude of Y should depend on the privacy loss parameter ε. A smaller value of ε is associated with a larger noise magnitude. When choosing the noise distribution, one possibility is to sample the random noise Y from a normal distribution with zero mean and standard deviation $1/\varepsilon$.[261] Because the choice of the value of ε is inversely related to the magnitude of the noise introduced by the analysis, the mechanism is designed to

260. *See supra* note 84 and accompanying text. The term "magnitude" refers to the magnitude of the random noise distribution as measured in parameters like the standard deviation or variance. This is not necessarily referring to the magnitude of the actual random noise sampled from the noise distribution. Generally, greater uncertainty requires a larger noise magnitude.

261. More accurately, the noise Y is sampled from the Laplace distribution with a mean of 0 and standard deviation of $\sqrt{2}/\varepsilon$. The exact shape of the noise distribution is important for proving that outputting $m + Y$ preserves differential privacy, but can be ignored for the current discussion.

provide a quantifiable tradeoff between privacy and utility.[262] Consider the following example.

Example 17

A researcher uses the estimate m', as defined in the previous example, to approximate the fraction p of HIV-positive people in the population. The computation would result in the estimate

$$p' = \frac{m'}{n} = \frac{38 + Y}{10{,}000}.$$

For instance, suppose the sampled noise is $Y = 4.2$. Then, the estimate would be

$$p' = \frac{38 + Y}{10{,}000} = \frac{38 + 4.2}{10{,}000} = \frac{42.2}{10{,}000} = 0.42\%,$$

whereas, without added noise, the estimate would have been $p = 0.38\%$.

A.2 Two Sources of Error: Sampling Error and Added Noise

This Section continues with the example from the previous Section. Note that there are two sources of error in estimating p: sampling error and added noise. The first source, sampling error, would cause m to differ from the expected $p \cdot n$ by an amount of roughly

$$|m - p \cdot n| \approx \sqrt{p \cdot n}.$$[263]

For instance, consider how the researcher from the example above would calculate the sampling error associated with her estimate.

262. Note that this means that, when the sample size is small, the accuracy can be significantly reduced. For instance, if the sample size is similar in magnitude to $1/\varepsilon$, the amount of noise that is added can even be larger than the sample size. Differential privacy works best when the sample size is large, specifically when it is significantly larger than $1/\varepsilon$.

263. The standard deviation of the difference $m - p \cdot n$ is $\sqrt{p \cdot (1-p) \cdot n} \approx \sqrt{p \cdot n}$ for small values of p. *See* BLITZSTEIN & HWANG, *supra* note 102, at 158–60. Thus, the expected value of the deviation $|m - p \cdot n|$ is approximately $\sqrt{p \cdot n}$. *See* J. Martin Bland & Douglas G. Altman, *Measuring Agreement in Method Comparison Studies*, 8 STAT. METHODS MED. RES. 135, 147 (1999).

Example 18

The researcher reasons that m' is expected to differ from $p \cdot 10{,}000$ by roughly

$$\sqrt{p \cdot 10{,}000} \approx \sqrt{38} \approx 6.$$

Hence, the estimate 0.38% is expected to differ from the true p by approximately

$$\frac{6}{10{,}000} = 0.06\%,$$

even prior to the addition of the noise Y by the differentially private mechanism.

The second source of error is the addition of random noise Y in order to achieve differential privacy. This noise would cause m' and m to differ by an amount of roughly

$$|m' - m| \approx 1/\varepsilon.$$ [264]

The researcher in the example would calculate this error as follows.

Example 19

The researcher reasons that, with a choice of $\varepsilon = 0.1$, she should expect $|m' - m| \approx 1/0.1 = 10$, which can shift p' from the true p by an additional $\frac{10}{10{,}000} = 0.1\%$.

Taking both sources of noise into account, the researcher calculates that the difference between noisy estimate p' and the true p is at most roughly

$$0.06\% + 0.1\% = 0.16\%.$$

264. The expectation of m' is exactly m because the Laplace distribution has zero mean. The standard deviation of the difference $m' - m$ is exactly the standard deviation of Y, which was chosen to be $1/\varepsilon$.

The two sources of noise are statistically independent,[265] so the researcher can use the fact that their variances add to produce a slightly better bound:

$$|p' - p| \approx \sqrt{0.06^2 + 0.1^2} = 0.12\%.$$

Generalizing from this example, we find that the standard deviation of the estimate p' (hence the expected difference between p' and p) is of magnitude roughly

$$|p' - p| \approx \sqrt{\frac{p}{n}} + \frac{1}{n\varepsilon}.$$

Notice that for a large enough sample size n, the noise added for privacy protection ($1/n\varepsilon$) will be much smaller than the sampling error ($\sqrt{\mathrm{p/n}}$), due to the difference between having n and $\sqrt{\mathrm{n}}$ in the denominator, and thus privacy comes essentially "for free" in this regime. Note also that the literature on differentially private algorithms has identified many other noise introduction techniques that can result in better accuracy guarantees than the simple technique used in the examples above.[266] Such techniques are especially important for more complex analyses, for which the simple noise addition technique discussed in this Section is often far from optimal in terms of accuracy.

A.3 Group Privacy

By holding individuals' opt-out scenarios as the relevant baseline, the definition of differential privacy directly addresses disclosures of information localized to a single individual. However, in many cases, information may be shared between multiple individuals. For example, relatives may share an address or certain genetic attributes.

How does differential privacy protect information of this nature? Consider the opt-out scenario for a group of k individuals. This is the scenario in which the personal information of all k individuals is omitted from the input to the analysis. For instance, John and Gertrude's opt-out scenario ($k = 2$) is the scenario in which both John's

265. Events are said to be statistically independent when the probability of occurrence of each event does not depend on whether the other event occurs. *See* BLITZSTEIN & HWANG, *supra* note 102, at 56.

266. *See* DWORK & ROTH, *supra* note 25, at 6, 22.

and Gertrude's information is omitted from the input to the analysis. Recall that the parameter ε controls how much the real-world scenario can differ from any individual's opt-out scenario. It can be shown that the difference between the differentially private real-world and opt-out scenarios of a group of k individuals grows to at most

$$k \cdot \varepsilon.$$ [267]

This means that the privacy guarantee degrades moderately as the size of the group increases. Effectively, a meaningful privacy guarantee can be provided to groups of individuals of a size of up to about

$$k \approx 1/\varepsilon$$

individuals.[268] However, almost no protection is guaranteed to groups of

$$k \approx 10/\varepsilon$$

individuals or greater.[269] This is the result of a design choice to not a priori prevent analysts using differentially private mechanisms from discovering trends across moderately-sized groups.[270]

267. *See id.* at 20; Dwork et al., *supra* note 62, at 29; Vadhan, *supra* note 46, at 361.

268. *See* DWORK & ROTH, *supra* note 25, at 192. When k is approximately $1/\varepsilon$, the group privacy guarantee corresponds to $k \cdot \varepsilon \approx 1$.

269. Guarantees that correspond to higher values than $k \cdot \varepsilon \approx 1$ (say, $k \cdot \varepsilon > 10$) provide only weak privacy guarantees.

270. *See generally* Dwork et al., *supra* note 62.