Bridging the Gap between Computer Science and Legal Approaches to Privacy

Kobbi Nissim[3,1], Aaron Bembenek[1], Alexandra Wood[2], Mark Bun[1], Marco Gaboardi[4], Urs Gasser[2], David R. O'Brien[2], Thomas Steinke[1], and Salil Vadhan[1]

[1]Center for Research on Computation and Society, Harvard University.
{tsteinke|mbun|salil}@seas.harvard.edu, bembenek@g.harvard.edu.
[2]Berkman Klein Center for Internet & Society, Harvard University.
{awood|ugasser|dobrien}@cyber.law.harvard.edu.
[3]Dept. of Computer Science, Georgetown University. kobbi.nissim@georgetown.edu
[4]The State University of New York at Buffalo. gaboardi@buffalo.edu.

February 21, 2018

This article is the product of a working group of the *Privacy Tools for Sharing Research Data* project at Harvard University.* Kobbi Nissim led the working group discussions that informed the development of the approach articulated in this article. Aaron Bembenek, Kobbi Nissim, and Alexandra Wood were the lead authors of the manuscript. Working group members Mark Bun, Marco Gaboardi, Urs Gasser, Kobbi Nissim, David O'Brien, Thomas Steinke, Salil Vadhan, and Alexandra Wood contributed to the conception of this work. The authors wish to thank Joseph Calandrino, Jonathan Frankle, Simson Garfinkel, Ann Kristin Glenster, Deborah Hurley, Georgios Kellaris, Paul Ohm, Leah Plunkett, Michel Reymond, Or Sheffet, Olga Slobodyanyuk, and the members of the *Privacy Tools for Sharing Research Data* project for their helpful comments. A preliminary version of this work was presented at the 9th Annual Privacy Law Scholars Conference (PLSC 2016), and the authors thank the participants for contributing thoughtful feedback. This material is based upon research supported by the National Science Foundation under Grant No. 1237235, grants from the Alfred P. Sloan Foundation, and a Simons Investigator grant to Salil Vadhan.

* Privacy Tools for Sharing Research Data, http://privacytools.seas.harvard.edu.

Abstract

The fields of law and computer science have generated different notions of privacy risks in the context of the analysis and release of statistical data about individuals. Emerging concepts from the theoretical computer science literature provide formal mathematical models for quantifying and mitigating privacy risks. Such models take into account a notion of privacy risk that is substantially broader than the privacy risks contemplated by many privacy laws. An example of a formal privacy model is *differential privacy*, which provides a concrete provable guarantee of privacy against a wide range of potential attacks, including types of attacks currently unknown or unforeseen. The subject of much theoretical investigation, new privacy technologies based on formal models such as differential privacy have recently been making significant strides towards practical implementation. For these tools to be used with sensitive personal information, it is important to demonstrate that they satisfy relevant legal requirements for privacy protection. However, making such an argument is challenging due to the significant conceptual gaps between the legal and technical approaches to defining privacy. Notably, information privacy laws are generally subject to interpretation and some degree of flexibility, which creates uncertainty for the implementation of more formal approaches.

This Article articulates the nature of the gaps between legal and technical approaches to privacy in the release of statistical data about individuals. It also presents an argument that the use of differential privacy is sufficient to satisfy the requirements of the Family Educational Rights and Privacy Act of 1974 (FERPA), a federal law that protects the privacy of education records in the United States. This argument illustrates what may evolve to a more general methodology for rigorously arguing that technological methods for privacy protection satisfy the requirements of a particular information privacy law. The argument detailed in this article has two main components. First, it involves the extraction of a formal mathematical requirement of privacy protection based on the standard set forth by FERPA. Second, it describes the construction of a rigorous mathematical proof for establishing that differential privacy satisfies the mathematical requirement extracted from FERPA. The argument takes a conservative "worst-case" approach in order to extract a mathematical requirement that is robust to potential ambiguities in legal interpretation. In this way, the mathematical proof demonstrates that the use of differential privacy is sufficient to satisfy a broad range of reasonable interpretations of FERPA, including interpretations that may be adopted in the future.

Keywords: privacy, privacy regulation, information privacy law, FERPA, de-identification, formal privacy models, differential privacy, game-based definitions

Contents

1 Introduction

The analysis and release of statistical data about individuals and groups of individuals carries inherent privacy risks,[1] and these risks have been conceptualized in different ways within the fields of law and computer science.[2] For instance, many information privacy laws adopt notions of privacy risk that are sector- or context-specific, such as in the case of laws that protect certain types of information contained within health, educational, or financial records from disclosure.[3] In addition, many privacy laws refer to specific techniques, such as de-identification, that are designed to address a subset of possible attacks on privacy.[4] In addition, many legal standards for privacy protection rely on individual organizations to make case-by-case determinations regarding concepts such as the identifiability of the types of information they hold. These regulatory approaches are intended to be flexible, allowing organizations to implement a variety of specific privacy measures that are appropriate given their varying institutional policies and needs, adapt to evolving best practices, and address a range of privacy-related harms. However, in the absence of clear thresholds and detailed guidance on making case-specific determinations, such flexibility in the interpretation and application of such standards also creates uncertainty for practitioners and often results in ad-hoc, heuristic processes. This uncertainty may pose a barrier to the adoption of new technologies that depend on unambiguous privacy requirements. It can, in turn, also lead organizations to implement measures that fall short of protecting against the full range of data privacy risks.

Emerging concepts from computer science provide formal mathematical models for quantifying and mitigating privacy risks that differ signficantly from traditional approaches to privacy such as de-identification. This creates conceptual challenges for the interpretation and application of existing legal standards, many of which implicitly or explicitly adopt concepts based on a de-identification approach to privacy. An example of a formal privacy model is *differential privacy*, which provides a provable guarantee of privacy against a wide range of potential attacks, including types of attacks currently unknown or unforeseen.[5] The subject of much theoretical investigation,

[1] *See generally* Cynthia Dwork, Adam Smith, Thomas Steinke, Jonathan Ullman, & Salil Vadhan, *Robust Traceability from Trace Amounts*, IEEE 56th Annual Symposium on Foundations of Computer Science (2015).

[2] For a discussion of the differences between legal and computer science definitions of privacy, see Felix T. Wu, *Defining Privacy and Utility in Data Sets*, 84 U. Colo. L. Rev. 1117 (2013).

[3] *See, e.g.,* Health Insurance Portability and Accountability Act Privacy Rule, 45 C.F.R. Part 160 and Subparts A and E of Part 164 (protecting "individually identifiable health information transmitted or maintained by a covered entity or its business associates in any form or medium"); Family Educational Rights and Privacy Act, 20 U.S.C. § 1232g, 34 C.F.R. Part 99 (protecting non-directory "personally identifiable information contained in education records"); Gramm-Leach-Bliley Act, 15 U.S.C. §§ 6801 et seq. (protecting "personally identifiable financial information provided by a consumer to a financial institution; resulting from any transaction with the consumer or any service performed for the consumer; or otherwise obtained by the financial institution").

[4] *See* Department of Health & Human Services Office of Civil Rights, *Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule*, https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/understanding/coveredentities/De-identification/hhs_deid_guidance.pdf (2012).

[5] Differential privacy was introduced in Cynthia Dwork, Frank McSherry, Kobbi Nissim & Adam Smith, *Calibrating Noise to Sensitivity in Private Data Analysis*, Proceedings of the 3rd Conference on the Theory of Cryptography 265 (2006). For a general introduction to differential privacy, see Kobbi Nissim, Thomas Steinke, Alexandra Wood, Micah Altman, Aaron Bembenek, Mark Bun, Marco Gaboardi, David O'Brien & Salil Vadhan, *Differential Privacy: A Primer for a Non-technical Audience (Preliminary Version)* (2017), http://privacytools.seas.harvard.edu/publications/differential-privacy-primer-non-technical-audience-preliminary-version (last updated May 16, 2017); Cynthia Dwork, *A Firm Foundation for Private Data Analysis*, 54 COMM. ACM 86 (2011); Ori Heffetz & Katrina Ligett, *Privacy and Data-Based Research*, 28 J. Econ. Persp. 75

new technologies relying on differential privacy are now making significant strides towards practical implementation. Several first-generation real-world implementations of differential privacy have been deployed by organizations such as Google, Apple, Uber, and the U.S. Census Bureau, and researchers in industry and academia are currently building and testing additional tools for differentially private statistical analysis.[6]

For these new technological tools to be used with sensitive personal information, being able to demonstrate that they satisfy relevant legal requirements for privacy protection will be key. However, making such an argument is challenging due to significant gaps between legal and mathematical approaches to defining privacy.[7] For instance, legal standards for privacy protection, and the definitions they employ, often vary according to industry sector, jurisdiction, institution, types of information involved, or other contextual factors.[8] Variations between laws create challenges for interpretation in the implementation of technological tools for privacy protection that are designed to be broadly-applicable. Legal approaches often, implicitly or explicitly, focus on a limited scope of attacks, such as re-identification by matching a named individual to a record in a database through linkage to information from other sources such as public records databases. This conceptualization of privacy risks, in part, leads many legal standards to turn on the presence of *personally identifiable information* in a data release. The concept of personally identifiable information is defined differently in various settings,[9] involves substantial ambiguity, and does not have a clear analog in mathematical definitions of privacy.[10]

In addition, standards and implementation guidance for privacy regulations such as the Family Educational Rights and Privacy Act of 1974 (FERPA),[11] Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule,[12] and the Privacy Act of 1974,[13] often emphasize techniques for protecting information released at the individual level but provide little guidance for releasing aggregate data, where the latter setting is particularly relevant to formal privacy models.[14] In

(2014); Erica Klarreich, *Privacy by the Numbers: A New Approach to Safeguarding Data*, Quanta Mag. (Dec. 10, 2012), https://www.quantamagazine.org/20121210-privacy-by-the-numbers-a-new-approach-to-safeguarding-data.

[6] *See* U.S. Census Bureau, OnTheMap, http://onthemap.ces.census.gov (last visited Jan. 22, 2016); Úlfar Erlingsson, Vasyl Pihur & Aleksandra Korolova, *RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response*, Proceedings of the 21st ACM Conference on Computer and Communications Security (2014); Andrew Eland, *Tackling Urban Mobility with Technology*, Google Europe Blog (Nov. 18, 2015), http://googlepolicyeurope.blogspot.com/2015/11/tackling-urban-mobility-with-technology.html; Microsoft Research, Privacy Integrated Queries (PINQ) Project, http://research.microsoft.com/en-us/projects/pinq (last visited Apr. 26, 2016); University of Pennsylvania, Putting Differential Privacy to Work Project, http://privacy.cis.upenn.edu/index.html (last visited Apr. 26, 2016); University of Texas, Airavat Project, http://z.cs.utexas.edu/users/osa/airavat (last visited Apr. 26, 2016); Prashanth Mohan et al., *GUPT: Privacy Preserving Data Analysis Made Easy*, Proceedings of SIGMOD '12 (2012); Privacy Tools for Sharing Research Data Project at Harvard University, Differentially Private Statistical Exploration, https://beta.dataverse.org/custom/DifferentialPrivacyPrototype (last visited Apr. 26, 2016).

[7] For an extended discussion of this argument, see Section 3.3 below.

[8] *See* 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 (2011).

[9] *See id.*

[10] *See* Arvind Narayanan & Vitaly Shmatikov, *Myths and Fallacies of "Personally Identifiable Information,"* 53 Communications of the ACM 24, 26 (2010).

[11] Family Educational Rights and Privacy Act of 1974, 20 U.S.C. § 1232g; 34 C.F.R. Part 99 (2013).

[12] Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, 45 C.F.R. Part 160 and Subparts A and E of Part 164

[13] 5 U.S.C. § 552a.

[14] Although documents have been produced to describe various techniques that are available for protecting privacy in aggregate data releases, *see, e.g.,* Federal Committee on Statistical Methodology, Re-

addition, limited or no guidance is available for cases in which personal information may be partly leaked, or in which it may be possible to make inferences about individuals' personal information with less than absolute certainty. Mathematical models of privacy must make determinations within difficult gray areas, where the boundaries of cognizable legal harms may not be fully defined. For these reasons, current regulatory terminology and concepts do not provide clear guidance for the implementation of formal privacy models such as differential privacy.

1.1 Contributions of this Article

The main contribution of this article is an argument that bridges between the definition of privacy in FERPA,[15] a federal law that protects the privacy of education records in the United States, and differential privacy,[16] a formal mathematical model of privacy. Two arguments are made along the way. The first is a legal argument supported by a technical argument that FERPA's requirements for privacy protection are relevant to analyses computed with differential privacy. The second is a technical argument supported by a legal argument that differential privacy satisfies FERPA's requirements for privacy protection. To address the inherent ambiguity that is reflected in different interpretations of FERPA, the analysis takes a conservative, worst-case approach and extracts a mathematical requirement that is robust to differences in interpretation. The resulting argument thereby demonstrates that differential privacy satisfies a large class of reasonable interpretations of the FERPA privacy standard.

While FERPA and differential privacy are used to illustrate an application of our approach, we believe the argument we present in this Article is an example of a more general approach. In future work, the approach we propose may be extended, with potential modifications, to bridge between technologies other than differential privacy and privacy laws other than FERPA. Its degree of rigor enables the formulation of strong arguments about the privacy requirements of statutes and regulations. Further, with the level of generalization afforded by this conservative approach to modeling, differences between sector- and institution-specific standards are blurred, making significant portions of the argument broadly-applicable. Future research exploring potential modifications to the argument and the extent to which it can address different interpretations of various legal standard could inform understandings of the significance of variations between different legal standards and, similarly, with respect to different technological standards of privacy.[17]

Arguments that are rigorous from both a legal and technical standpoint can help support the future adoption of emerging privacy-preserving technologies. Such arguments could be used to assure actors who release data with the protections afforded by a particular privacy technology that they are doing so in compliance with the law. They could also be used to better inform data subjects of the privacy protection to which they are legally entitled and to demonstrate that

port on Statistical Disclosure Limitation Methodology, Statistical Policy Working Paper 22 (2005), http://www.hhs.gov/sites/default/files/spwp22.pdf, such documents are not designed to help practitioners determine when the use of certain techniques is sufficient to meet regulatory requirements. For an extended discussion of the laws governing privacy in government data releases and the lack of guidance available for selecting among the broad range of privacy techniques and applying tools in specific cases, see Micah Altman, Alexandra Wood, David R. O'Brien, Salil Vadhan & Urs Gasser, *Towards a Modern Approach to Privacy-Aware Government Data Releases*, 30 Berkeley Tech. L.J. 1967, 1975–2009 (2015).

[15] Family Educational Rights and Privacy Act of 1974, 20 U.S.C. § 1232g; 34 C.F.R. Part 99 (2013).

[16] See *supra* note 5.

[17] For a further discussion of anticipated differences with respect to modeling legal standards beyond FERPA, see Section 5.

their personal information will be handled in accordance with that guarantee of privacy protection. More generally, the process of formalizing privacy statutes and regulations can lead scholars and practitioners to better understand their requirements and help identify aspects of law and policy that are ambiguous or insufficient to address real-world privacy risks. In turn, it can serve as a foundation for future extensions to address other problems in information privacy, an area that is both highly technical and highly regulated and therefore is well-suited to combined legal-technical solutions.[18]

Note that, while this Article demonstrates that differential privacy is sufficient to satisfy the requirements for privacy protection set forth by FERPA, it does not make the claim that differential privacy is the only technological standard that could be shown to be sufficient under the law. Indeed, we expect that extensions analyzing other technologies would support sufficiency arguments with respect to various legal and policy requirements for privacy protection. Differential privacy is just one of a large collection of privacy interventions that may be considered as an appropriate component of an organization's data management program. This Article is intended to complement other resources providing practical guidance on incorporating data sharing models, including those that rely on formal privacy frameworks like differential privacy, within a specific institutional setting.[19]

1.2 Article structure

In the sections that follow, we establish a rigorous approach to modeling FERPA and bridging between differential privacy and FERPA's privacy requirements. Section 2 describes the setting in which the privacy issues relevant to this discussion arise, and provides a high-level overview of approaches from computer science and law that have emerged to address these issues. Section 3 provides an introduction to the two privacy concepts at the heart of our analysis, differential privacy and FERPA. It also discusses the applicability of FERPA's requirements to differentially private computations and articulates the gaps between differential privacy and FERPA that create challenges for implementing differential privacy in practice. The later sections present a novel argument for formally proving that differential privacy satisfies FERPA's privacy requirements. Section 4 describes the process of extracting a formal mathematical requirement of privacy protection under FERPA. It demonstrates how to construct a model of the attacker that is implicitly recognized by FERPA, based on the definitions found in the FERPA regulations and informed by the regulatory history and related administrative guidance. Section 5 concludes with a discussion exploring how formal argumentation can help enable real-world implementation of emerging formal privacy technologies and the development of more robust privacy regulations.

[18] For other research formalizing legal requirements for privacy protection using mathematical approaches, see, e.g., Omar Chowdhury, Andreas Gampe, Jianwei Niu, Jeffery von Ronne, Jared Bennatt, Anupam Datta & Limin Jia, *Privacy Promises That Can Be Kept: A Policy Analysis Method with Application to the HIPAA Privacy Rule*, Proceedings of the 18th ACM Symposium on Access Control Models and Technologies 3 (2013); Henry DeYoung, Deepak Garg, Dilsun Kaynar, & Anupam Datta, *Logical Specification of the GLBA and HIPAA Privacy Laws*, Proceedings of 9th ACM Workshop on Privacy in the Electronic Society (2010).

[19] For an extended discussion of the array of legal, technical, and procedural privacy controls that are available and how to determine which are suitably aligned with the privacy risks and intended uses in a particular context, see Micah Altman, Alexandra Wood, David R. O'Brien, Salil Vadhan & Urs Gasser, *Towards a Modern Approach to Privacy-Aware Government Data Releases*, 30 Berkeley Tech. L.J. 1967, 1975–2009 (2015). *See also* Simson Garfinkel, *De-identifying Government Datasets*, NIST Special Publication 800-188 (2nd Draft) (2016), http://csrc.nist.gov/publications/drafts/800-188/sp800_188_draft2.pdf.

For reference, we also include two Appendices with supplementary arguments and models. Appendix A provides a sketch of the mathematical proof to demonstrate that differential privacy satisfies the mathematical definition of privacy extracted from FERPA. Appendix B presents two possible extensions of the model described in Section 4.

2 Background

Privacy is conceptualized differently across a range of legal contexts, from surveillance to criminal procedure to public records releases to research ethics to medical decision making.[20] Because privacy law is quite broad in its reach and privacy measures can be designed to address a wide range of harms, it is important to define the scope of the analysis in this Article. Specifically, this Article focuses on privacy in the context of the statistical analysis of data or the release of statistics derived from personal data. We refer to the relevant setting as *privacy in statistical computation*.

2.1 The setting: Privacy in statistical computation

Many government agencies, commercial entities, and research organizations collect, process, and analyze data about individuals and groups of individuals. They also frequently release such data, or statistics based on the data. If the data being released contain personal information, various laws and policies likely restrict the degree to which the data can be disclosed, including the formats in which the data can be published.

Federal and state statistical agencies such as the Census Bureau, the Bureau of Labor Statistics, and the National Center for Education Statistics release large quantities of statistics about individuals, households, and establishments, upon which policy and business investment decisions are based. For instance, the National Center for Education Statistics collects data from U.S. schools and colleges, using surveys and administrative records, and releases statistics derived from these data to the public.[21] These statistics include total enrollments at public and private schools by grade level, class sizes, participation levels in activities such as reading at home, access to computers and technology, standardized test scores by state and student demographics, and high school graduation rates, among other figures, and are used to shape education policy and school practices.[22] The release of such statistics by federal statistical agencies is subject to strict requirements for protecting the privacy of the individuals in the data.[23]

[20] For a broad survey of privacy concepts, see, e.g., Daniel J. Solove, *A Taxonomy of Privacy*, 154 U. Penn. L. Rev. 477 (2006) (grouping privacy problems into categories such as surveillance, interrogation, aggregation, identification, insecurity, secondary use, exclusion, breach of confidentiality, disclosure, exposure, increased accessibility, blackmail, appropriation, distortion, intrusion, and decisional interference); Bert-Jaap Koops, Bryce Clayton Newell, Tjerk Timan, Ivan Škorvánek, Tomislav Chokrevski, & Maša Galič, *A Typology of Privacy*, 38 U. Pa. J. Int'l L. (forthcoming 2016) (surveying constitutional protections and privacy scholarship from nine North American and European countries, resulting in a classification of privacy based on "eight basic types of privacy (bodily, intellectual, spatial, decisional, communicational, associational, proprietary, and behavioral privacy), with an overlay of a ninth type (informational privacy) that overlaps, but does not coincide, with the eight basic types"); Alan F. Westin, Privacy and Freedom (1967) (identifying four states of privacy: solitude, intimacy, anonymity, and reserve).

[21] *See* National Center for Education Statistics, Digest of Education Statistics: 2013, NCES Pub. No. 2015-011 (2015).

[22] *See id.*

[23] *See* Confidential Information Protection and Statistical Efficiency Act, 44 U.S.C. § 3501 note (prohibiting the release of statistics in "identifiable form" and establishing specific confidentiality procedures to be followed by employees and contractors when collecting, processing, analyzing, and releasing data).

Companies such as Google and Facebook also collect personal information, which they use to provide services to individual users and third-party advertisers. For instance, Facebook enables advertisers to target ads on the Facebook platform based on the locations, demographics, interests, and behaviors of their target audiences, and provides tailored estimates of the number of recipients of an ad given different parameters.[24] The commercial collection, use, and release of data is regulated by the Federal Trade Commission,[25] as well as any applicable federal and state information privacy laws.[26]

Researchers and their institutions also release statistics to the public and make their data available to other researchers for secondary analysis. These research activities are subject to oversight by an institutional review board in accordance with the Federal Policy for the Protection of Human Subjects.[27] Researchers governed by these regulations are required to obtain consent from participants and to employ data privacy safeguards when collecting, storing, and sharing data about individuals.[28]

In the discussion that follows, we explore how statistical computations can be performed by government agencies, commercial entities, and researchers while providing strong privacy protection to individuals in the data.

2.1.1 What is a computation?

A *computation* (alternatively referred to as an algorithm, mechanism, or analysis) is a mechanizable procedure for producing an output given some input data, as illustrated in Figure 1.[29] This general definition does not restrict the nature of the relationship between the input data and the output. For instance, a computation could output its input without any transformation, or it could even ignore its input and produce an output that is independent of the input. Some computations are *deterministic* and others are *randomized*. A deterministic computation will always produce the same output given the same input, while a randomized computation does not provide such a guarantee. A computation that returns the mean age of participants in a dataset is an example of a deterministic computation. However, a similar computation that estimates the mean age in the dataset by sampling at random a small subset of the records in a database and returning the mean age computed for these records is a randomized computation. As another example, a computation that outputs the mean age with the addition of some random noise is a randomized computation.[30]

Public releases of statistics derived from personal information, such as the statistical information many government agencies publish, result from applying certain computations on the personal

[24] *See* Facebook, Facebook for Business: How to target Facebook ads, https://www.facebook.com/business/a/online-sales/ad-targeting-details (last visited Apr. 27, 2016). These audience-reach statistics are evidently rounded, in part, to protect the privacy of Facebook users. *See* Andrew Chin & Anne Klinefelter, *Differential Privacy as a Response to the Reidentification Threat: The Facebook Advertiser Case Study*, 90 N. Carolina L. Rev. 1417 (2012).

[25] *See* Federal Trade Commission, Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers, FTC Report (2012).

[26] *See, e.g.,* Children's Online Privacy Protection Act, 15 U.S.C. §§ 6501–6506 (requiring certain web site operators to provide notice and obtain consent when collecting personal information from children under 13 years of age); Cal. Civ. Code § 1798.82 (requiring businesses to disclose any data breach to California residents whose unencrypted personal information was acquired by an unauthorized person).

[27] *See* 45 C.F.R. Part 46.

[28] 45 C.F.R. § 46.111.

[29] This figure is reproduced from Kobbi Nissim, Thomas Steinke, Alexandra Wood, Micah Altman, Aaron Bembenek, Mark Bun, Marco Gaboardi, David O'Brien & Salil Vadhan, *Differential Privacy: A Primer for a Non-technical*

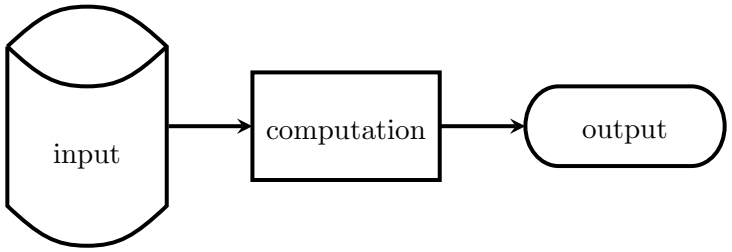

Figure 1: A computation produces an output given some input data.

information collected. The U.S. Census Bureau, for example, collects information about individual households, including sensitive personal information, and then releases statistics about the larger population. Statistics such as the median household income of a city or region are computed after collecting household-level income information from a sample population in that area. Similarly, a school district, seeking to study and report on the performance of students in its schools, applies a set of computations to the raw data it has collected about individual students to generate more general statistics for publication. Releasing statistics about personal information can benefit society by enabling research and informing policy decisions. However, the computations used to create the statistics must provide sufficient privacy protection for the individuals in the data. In the next subsection, we discuss how it is possible to release statistics based on personal information while protecting individual privacy.

2.1.2 Privacy as a property of computation

This Article considers whether a computation provides privacy protection to be a property of that computation. To see why this approach makes sense from an informational point of view, note that privacy risks are created when the output of a computation carries information about its input. Consider, for example, a computation that receives as input a student's transcript and outputs the student's GPA. Seeing that the outcome of this computation is a GPA of 3.7, an observer can rule out certain transcripts for the student that would result in a different GPA. In some cases, the observer would be able to uniquely determine the student's grades (e.g., when the GPA is zero, or when the observer has prior information that, together with the computed GPA, uniquely determines the grades).[31]

When we make the claim that a certain computation provides privacy protection, we mean that the *informational relationship between input and output* of this computation satisfies the requirements of a particular definition of privacy. We emphasize that it is the computation that is private rather than a particular output that is private. To see why distinguishing between a purportedly

Audience (Preliminary Version) (2017), in which an extended discussion of privacy in computation can also be found.

[30] Randomness may be introduced into computations for a number of reasons. Foremost among them are efficiency, and the need to create uncertainty for an adversary (e.g., when cryptographic keys are selected or for achieving privacy protection).

[31] While this example is using a deterministic computation of the GPA for simplicity, it can be also generalized to cases of randomized computations. That is to say that, while some randomized algorithms introduce uncertainty in ways that effectively mitigate leakage of sensitive information, there also exist randomized algorithms that can result in significant leakages, similar to the example of a deterministic GPA computation.

private or *non-private* output fails to capture a reasonable notion of privacy, consider a policy that states that statistics needs to be coarsened to the nearest ten (e.g., 0–9, 10–19, 20–29, etc.), with the hope that such a coarsening would hide the effect of individuals. Suppose a school releases statistics for the fall semester showing that 20–29 of its students have a disability, an output that may seem innocuous in terms of risk to individual privacy. In the spring, the school releases updated statistics showing that 30–39 of its students have a disability, another output that may seem innocuous. However, reasoning about how the two statistics were computed reveals that a new student who joined between the fall and spring semesters has a disability. Although both *outputs* seem innocuous as each output only reveals aggregate statistics and does not directly identify any individual student, reasoning about how they *depend on the input data*—a dependency created by the computation of the statistics—reveals sensitive information.

Computer scientists seek to reason about the properties of computations, and treating privacy as a computational property fits naturally in this world view. This approach has successful precedents in established areas of computer science such as cryptography.[32] We argue that this approach—defining privacy as a property of a computation—is also applicable from a legal perspective. While privacy laws and policies might not explicitly refer to computation, we observe that they often attempt to implicitly define privacy as a property that certain computations possess. For instance, consider the HIPAA Privacy Rule's safe harbor method of de-identification, which is satisfied when certain pieces of personal information deemed to be identifying have been removed from a dataset prior to release.[33] This provision is in effect specifying a computation. A computation that produces an output in which the enumerated identifiers have been redacted is considered to provide sufficient privacy protection. Similarly, regulatory requirements or related guidance prescribing minimum cell counts for aggregate data tables produced using personal information,[34] or recommending transforming data using k-anonymization techniques prior to release,[35] are effectively treating certain computations as providing privacy protection.[36]

[32] For instance, a cryptographic computation might be considered to provide privacy protection if any hypothetical adversary given an encrypted message can do no better at guessing a property of that message than another hypothetical adversary that is not given the encrypted message. *See* Shafi Goldwasser & Silvio Micali, *Probabilistic Encryption and How to Play Mental Poker Keeping Secret All Partial Information*, Proceedings of ACM Symposium on Theory of Computing 36 (1982).

[33] 45 C.F.R. § 164.514

[34] For example, in accordance with the Elementary and Secondary Education Act of 1965, states must define minimum cell counts for the publication of student achievement results "[b]ased on sound statistical methodology" that "[y]ields statistically reliable information for each purpose for which disaggregated data are used" and does not "reveal personally identifiable information about an individual student." 34 C.F.R. § 200.7.

[35] "A release provides k-anonymity protection if the information for each person contained in the release cannot be distinguished from at least $k-1$ individuals whose information also appears in the release." *See* Latanya Sweeney, *k-Anonymity: A Model for Protecting Privacy*, 10 International Journal on Uncertainty, Fuzziness and Knowledge-based Systems 557 (2002). Guidance from the Department of Health and Human Services covers the application of k-anonymity as one approach to protectng health records subject to the HIPAA Privacy Rule. *See* Office for Civil Rights, Department of Health and Human Services, *Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule* (2012) (providing guidance on applying k-anonymity and noting that "[t]he value for k should be set at a level that is appropriate to mitigate risk of identification by the anticipated recipient of the data set," while declining to "designate a universal value for k that covered entities should apply to protect health information in accordance with the de-identification standard").

[36] Note that by specifying properties of the output, such requirements restrict the computation, but do not necessarily successfully restrict the informational relationship between input and output. We include these examples here only to support the claim that conceptualizing the problem in terms of restrictions on computation is compatible

A benefit of viewing privacy as a property of computation is that computations are formal mathematical objects, and as such can be reasoned about with a high degree of mathematical rigor. In this Article we attempt to use this view to make rigorous arguments about the privacy requirements of laws and regulations. In turn, these arguments can assure actors who release data that they are following the law, and may give data subjects a better understanding of the privacy protection to which they are legally entitled. Additionally, the process itself of formalizing legal requirements as computational objects can lead us to better understand those requirements and help identify aspects of the law that are potentially ambiguous or may be viewed as insufficient to provide adequate privacy protection.

2.1.3 Privacy risks in computation

A number of approaches have been developed and widely implemented to limit the disclosure of personal information when sharing statistical data about individuals. Traditional approaches include obtaining consent from data subjects, entering into data use agreements restricting the use and re-disclosure of data, and applying various techniques for de-identifying data prior to release.[37] Statistics about individuals or groups of individuals are generally made available after de-identification techniques have transformed the data by removing, generalizing, aggregating, and adding noise to pieces of information determined to be identifiable. At the core of this approach is a concept of *personally identifiable information* (PII), which is based in a belief that privacy risks lurk in tables of individual-level information, and that protecting privacy depends on the removal of information deemed to be identifying, such as names, addresses, and Social Security numbers.[38]

Privacy in the release of statistical data about individuals is under increasing scrutiny. Numerous data privacy breaches have demonstrated that risks can be discovered even in releases of data that have been redacted of personally identifiable information. For example, in the late 1990s, Latanya Sweeney famously demonstrated that the medical record of Massachusetts Governor William Weld could be identified in a release of data on state employee hospital visits that had been stripped

with some existing legal and regulatory approaches to privacy protection.

[37] *See, e.g.,* Department of Education, Data-Sharing Tool Kit for Communities: How to Leverage Community Relationships While Protecting Student Privacy (March 2016), http://www2.ed.gov/programs/promiseneighborhoods/datasharingtool.pdf (providing best practice guidance on sharing education data while protecting privacy, by de-identifying data, obtaining written consent, or entering into a written data-sharing agreement with the recipient).

[38] Many privacy laws explicitly or implicitly endorse the practice of removing personal information considered to be identifying prior to release. *See, e.g,* Family Educational Rights and Privacy Act of 1974, 20 U.S.C. § 1232g; 34 C.F.R. Part 99 (permitting educational agencies and institutions to broadly release, without the consent of students or their parents, information from education records that have been de-identified through the removal of "personally identifiable information"); Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, 45 C.F.R. Part 160 and Subparts A and E of Part 164 (providing a safe harbor permitting the release of health information that has been de-identified through the removal of information from a list of eighteen "identifiers"). For an extended discussion of various legal approaches to de-identification, see Section 2.3 below. Such approaches also appear in a wide range of guidance materials on privacy and confidentiality in data management. *See, e.g.,* Federal Committee on Statistical Methodology, Report on Statistical Disclosure Limitation Methodology, Statistical Policy Working Paper 22 (2005), http://www.hhs.gov/sites/default/files/spwp22.pdf (recommending the removal of identifiers but also acknowledging that privacy risks may remain despite redaction: "After direct identifiers have been removed, a file may still remain identifiable, if sufficient data are left on the file with which to match with information from an external source that also contains names or other direct identifiers. For this reason, agencies should perform reidentification studies and attempt to match variables on the released files to external files outside of the agency.").

of the names and addresses of patients in order to protect their privacy.[39] Using Governor Weld's date of birth, ZIP code, and gender, which could be found in public records, she was able to locate his record in the dataset that had been released, as it was the only record that matched all three attributes. Indeed, well over 50% of the U.S. population can be uniquely identified using these three pieces of information.[40]

Repeated demonstrations across many types of data have confirmed that this type of privacy breach is not merely anecdotal but is in fact widespread. For instance, it has been shown that individuals can be identified in releases of Netflix viewing records and AOL search query histories, despite efforts to remove identifying information from the data prior to release.[41] Other research has demonstrated that just four records of an individual's location at a point in time can be sufficient to identify 95% of individuals in mobile phone data and 90% of individuals in credit card purchase data.[42] Narayanan and Shmatikov generalize these and other privacy failures stating that "[a]ny information that distinguishes one person from another can be used for re-identifying anonymous data."[43]

Other privacy attacks have exposed vulnerabilities in releases of aggregate data and revealed inference risks that are distinct from the risk of re-identification.[44] More specifically, many successful attacks on privacy have focused not on discovering the identities of individuals but rather on learning or inferring sensitive details about them. For example, researchers have discovered privacy risks in databases containing information about mixtures of genomic DNA from hundreds of people.[45] Although the data were believed to be sufficiently aggregated so as to pose little risk to the privacy of individuals, it was shown that an individual's participation in a study could be confirmed using the aggregated data, thereby revealing that the individual suffers from the medical condition that was the focus of the research study.

Privacy risks have also been uncovered in online recommendation systems used by web sites such as Amazon, Netflix, and Last.fm, which employ algorithms that recommend similar products

[39] *See* Latanya Sweeney, *Weaving Technology and Policy Together to Maintain Confidentiality*, 25 J. L., Med., & Ethics 98 (1997).

[40] *See* Latanya Sweeney, *Simple Demographics Often Identify People Uniquely*, Data Privacy Lab Technical Report (2000); Philippe Golle, *Revisiting the uniqueness of simple demographics in the US population*, Proceedings of the 2006 ACM Workshop on Privacy in the Electronic Society (WPES) 77 (2006).

[41] *See* Arvind Narayanan & Vitaly Shmatikov, *Robust De-anonymization of Large Sparse Datasets*, Proceedings of the 2008 IEEE Symposium on Research in Security and Privacy 111 (2008); Michael Barbaro & Tom Zeller, Jr., *A Face Is Exposed for AOL Searcher No. 4417749*, N.Y. Times, Aug. 9, 2006.

[42] *See* Yves-Alexandre de Montjoye et al., *Unique in the Shopping Mall: On the Reidentifiability of Credit Card Metadata*, 347 Science 536 (2015); Yves-Alexandre de Montjoye et al., *Unique in the Crowd: The Privacy Bounds of Human Mobility*, 3 Nature Sci. Reps. 1376 (2013).

[43] *See* Arvind Narayanan & Vitaly Shmatikov, *Myths and Fallacies of "Personally Identifiable Information,"* 53 Communications of the ACM 24, 26 (2010).

[44] For a recent survey of different classes of privacy attacks, including both re-identification attacks and tracing attacks (i.e., attacks that aim to determine whether information about a target individual is in a database), see the discussion in Cynthia Dwork, Adam Smith, Thomas Steinke & Jonathan Ullman, *Hiding in Plain Sight: A Survey of Attacks on Private Data* (forthcoming 2017).

[45] *See* Nils Homer et al., *Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-density SNP Genotyping Microarrays*, 4 PLoS Genetics 8 (2008). This attack involves using published genomic statistics and some information about the underlying population to determine whether an individual's genomic sequence was used in the study. This attack has been strengthened and generalized in several works, such as Cynthia Dwork et al., *Robust Traceability from Trace Amounts*, IEEE Symposium on Foundations of Computer Science (2015), http://privacytools.seas.harvard.edu/files/robust.pdf.

to users based on an analysis of data generated by the behavior of millions of users.[46] These attacks have shown that recommendation systems can leak information about the transactions made by individuals. In addition to such attacks that have been demonstrated on real-world data, theoretical work has resulted in other, more general privacy attacks on aggregate data releases a variety of settings.[47] These findings, taken together, demonstrate that privacy risks may be present when releasing not just individual-level records but also aggregate statistics, and that privacy risks include not only re-identification risks but other inference risks as well.

Moreover, techniques for inferring information about individuals in a data release are rapidly advancing and exposing vulnerabilities in many commonly used measures for protecting privacy. Although traditional approaches to privacy—including those that are referenced implicitly or explicitly by the de-identification requirements of information privacy laws—may have been sufficient at the time they were developed, they are becoming increasingly ill-suited for protecting information in the Internet age.[48] It is likely that privacy risks will continue to grow and evolve over time, enabled by rapid advances in analytical capabilities and the increasing availability of personal data from different sources, and motivated by the tremendous value of sensitive personal data. These realizations have led computer scientists to seek new approaches to data privacy that are robust against a wide range of attacks, including types of attacks unforeseen at the time they were developed.

2.2 An introduction to the computer science approach to defining privacy

In the field of computer science, privacy is often formalized as a game, or a thought experiment, in which an adversary attempts to exploit a computational system to learn protected information. A system is considered to provide privacy protection if it can be demonstrated, via a mathematical proof, that no adversary can win the game with a probability that is "too high." Every aspect of a game framework must be carefully and formally defined, including any constraints on the adversary, the mechanics of the game, what it means for the adversary to win the game, and with what probability it is acceptable for the adversary to win the game. This formalization allows us to prove that a given system is private under the explicit assumptions of the model.

The formal privacy games we discuss have the following components: an adversary, a computation, and game mechanics. Each of these components is discussed in turn below.

An *adversary* attempts to exploit the system to learn private information. The adversary is not defined in terms of the specific techniques he might use to exploit the system, but rather by the computational power, access to the system, and background knowledge he can leverage. As a consequence, the adversary does not represent a uniquely specified attack, but rather a whole

[46] *See* Joseph A. Calandrino et al., *"You Might Also Like:" Privacy Risks of Collaborative Filtering*, IEEE Symposium on Security and Privacy (2011).

[47] For example, Dinur and Nissim showed that publishing estimates to randomly issued statistical queries can lead to a very accurate reconstruction of the information in a statistical database, and hence result in a massive leakage of individual information. *See* Irit Dinur & Kobbi Nissim, *Revealing Information While Preserving Privacy*, Proceedings of the 22nd ACM PODS 202 (2003).

[48] It is important to also note that, in addition to vulnerability to privacy attacks, traditional de-identification techniques may be unsuitable in practice for other reasons, such as the impact their use has on data quality. *See, e.g.,* Jon P. Daries et al., *Privacy, Anonymity, and Big Data in the Social Sciences*, 12 ACM Queue (2014), http://queue.acm.org/detail.cfm?id=2661641 (In discussing the significantly different results in analyses based on massive open online course data before and after applying de-identification techniques, the authors note that "the original analysis found that approximately 5 percent of course registrants earned certificates. Some methods of de-identification cut that percentage in half.").

class of attacks, *including attacks that have not been conceived by the system designer*. This means that a system cannot be *tested* for its privacy, as testing its resilience to known attacks would not rule out its vulnerability to other attacks. Rather, privacy needs to be *proved* mathematically. By proving that the system provides privacy protection against such an adversary, we have made a strong claim: the system is private no matter the particular attack or attacker, provided that the assumptions we have made in our model are not violated.

A *computation* takes as input a dataset (potentially containing private information) and produces some output. For instance, one could envision a computation that takes as input a spreadsheet of students' grades and returns the average grade (or an approximation of it). Unlike the adversary, the computation needs to be uniquely specified[49] and, furthermore, known to the adversary.[50] We use the game framework to prove that a computation (or a class of computations) provides privacy protection given the assumptions of the game. For example, we might prove that in a particular game framework the computation reports the average grade in such a way as to maintain the individual privacy of the students in the original dataset.

The *game mechanics* act as an intermediary between the adversary and the private data.[51] The adversary does not have direct access to non-public data, and instead receives information via the game mechanics. The game mechanics enforce a specific protocol and determine the ways in which the adversary can interact with the computation. In addition, the game mechanics define when the adversary is deemed to have won the game and when a system is deemed to provide a sufficient level of privacy. For instance, the game mechanics might specify the following protocol: the previously described computation is used to calculate the average grade on a test, and then the computation result is released to the adversary. The adversary responds with a guess of the grade of a particular student, and is considered to have won the game if his guess is within a certain range of that student's true grade.

A privacy game provides us with a precise privacy definition. A computation satisfies the definition of privacy if no adversary can win the game (with the computation) "too often." Note that we do not require that an adversary never win the game. Such a requirement, even if intuitive, would not be achievable, as there is always some probability that an adversary could win the game by sheer chance, even without seeing any outcome of the computation. In other words, the standard is that no adversary should be able to win the game with a probability that is significantly greater than some baseline probability, which we would generally take to be the probability of winning without access to the computation outcome.

To illustrate, consider a game in which an adversary's goal is to guess a person's gender. It may be possible for the adversary to guess the person's gender correctly with the probability of approximately 50%, even without obtaining any information about that person. This means that the baseline for comparison should be a probability of (at least) 50% for guessing a target individual's gender, as it would not be reasonable to establish a requirement that all adversaries

[49] When we use the game framework to prove that a *class* of computations provides privacy, the proof assumes that an arbitrary computation of this class is selected as the uniquely specified computation before the execution of the game begins.

[50] Here we follow a design principle widely accepted in cryptography and known as *Kerckhoffs' principle*: a cryptographic system should maintain security even if all details of its design and implementation are known to the attacker. *See* Auguste Kerckhoffs, *La Cryptographie Militaire*, IX Journal des Sciences Militaire (1883).

[51] Note that the game mechanics does not necessarily correspond to any specific "real world" entity. A privacy game is a thought experiment and in general there is not a direct correspondence between the components of a privacy game and the parties in a real-world privacy scenario.

must guess correctly with a success rate that is lower than this probability. Additionally, we typically allow for adversaries to win with a probability that is slightly higher than the baseline value (because any system that provides utility necessarily leaks at least a tiny bit of information).[52] How much the adversary is allowed to win beyond the baseline probability while the computation is still considered to satisfy the privacy definition is often quantified as a parameter that can be tuned to provide less or more privacy protection.[53]

For a more detailed explanation, consider the following privacy game from cryptography. Alice wants to send a private message to Bob, but she is worried that a third party, Eve, might be eavesdropping on their communications. Therefore, Alice decides to encrypt the message before sending it to Bob. She wants to be confident that the computation she uses to encrypt the message will ensure that Eve cannot learn much about the content of the original message (i.e., the plaintext) from seeing the encrypted version of the message (i.e., the ciphertext). To gain confidence in the security of the system, Alice can formalize her privacy desiderata as a game and then use an encryption computation that is proven to meet the game's definition of privacy. Here is one possible description of the mechanics for the game, also represented visually in Figure 2:[54]

1. An adversary Eve chooses two distinct plaintext messages and passes them to the game mechanics. Intuitively, she is asserting that she can distinguish between the encrypted messages and hence the encryption is insecure.

2. The game mechanics tosses a fair coin to choose between the two messages with equal probability, encrypts the chosen message (denoted "plaintext" in Figure 2) with a computation C, and gives the resulting ciphertext to the adversary.

3. The adversary wins if she is able to guess from seeing the ciphertext which of the two original messages was encrypted.

Notice that an adversary that ignores the ciphertext she is given in Step 2 of the game and simply outputs one of the messages she selected in Step 1 already has a 50% chance of winning.

[52] Federal Committee on Statistical Methodology, Report on Statistical Disclosure Limitation Methodology, Statistical Policy Working Paper 22 (2005) states "The release of statistical data inevitably reveals some information about individual data subjects." Intuitively, a release "provides utility" only if it reveals some information that was not previously known about the group of individuals whose information is in the dataset. In information theory this is referred to as a reduction in entropy, a quantity (measured in bits) that corresponds to the uncertainty an attacker may have about the data. Informally, if entropy is reduced by $k > 0$ bits and the data concerns n individuals, then there would be an individual that suffers entropy reduction of at least k/n bits. For a concrete example illustrating why absolute disclosure prevention is impossible in any system that provides utility, we consider an illustration by Dwork and Naor. Rephrasing their example, consider a privacy attacker who has access to the auxiliary information "Terry Gross is two inches shorter than the average American woman" but possesses no other knowledge about the height of American women (hence, the auxiliary information is initially not helpful to the attacker). Given access to a system that allows estimating the average height of American women, the attacker can estimate Terry Gross' height. Informally, the attacker's auxiliary information contained sensitive information (Terry Gross' height), but this information was "encrypted" and hence useless to the attacker. The decryption key (the average height of American women) could be learned by invoking the system's computation. *See* Cynthia Dwork and Moni Naor, *On the Difficulties of Disclosure Prevention in Statistical Databases or The Case for Differential Privacy*, Journal of privacy and confidentiality, Vol. 2 (2007-2011), Iss. 1 (2008-2010).

[53] For further discussion of the privacy parameter, see Section 4 below.

[54] This game is modeled after the notion of indistinguishability of ciphertexts introduced in Shafi Goldwasser & Silvio Micali, *Probabilistic Encryption and How to Play Mental Poker Keeping Secret All Partial Information*, Proceedings of ACM Symposium on Theory of Computing 36 (1982).

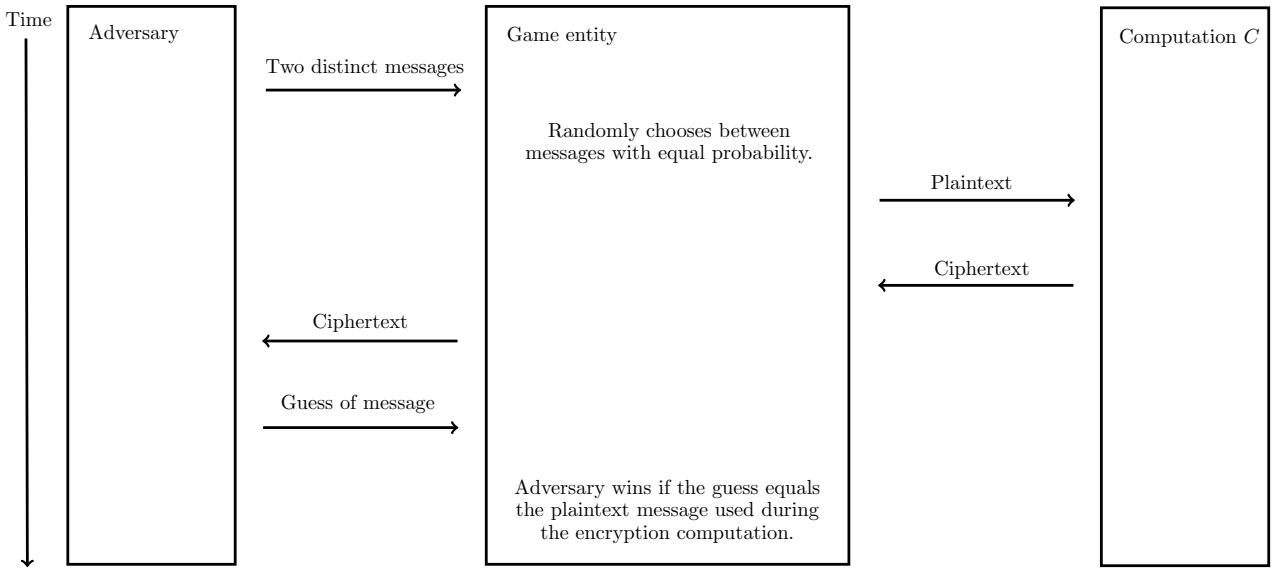

Figure 2: Example cryptographic privacy game.

This means that any reasonable adversary would have a winning probability of at least 50%, and we take the 50% success rate as a baseline for comparison. Therefore, a computation C for which no adversary can win the game with a probability greater than 50%, equal to the baseline, can be considered as providing perfect privacy protection. We typically relax this requirement to allow adversaries to obtain a small advantage over the 50% baseline. That is, a computation C would be considered to satisfy the privacy definition if no adversary can win this game with a probability significantly higher than 50%. The cryptographic standard for encryption is to only allow adversaries a negligible advantage over the 50% probability, say a 50.00000000001% chance of winning the game.[55] Now that her privacy desiderata have been formalized, Alice can use any encryption computation that is mathematically proven to meet this definition of privacy to send an encrypted message to Bob with confidence that Eve cannot learn much from eavesdropping on their communication.

Although privacy laws are not written with an explicit game-based definition of privacy, we argue that it is possible to extract a suitable privacy game from a law, its legislative history, and administrative guidance. Furthermore, the privacy game that is extracted can be used to establish that particular computations meet the privacy requirements of the law. In Section 4 below, we extract such a privacy game from FERPA. Based on this game, we sketch in Appendix A a proof showing that all differentially private computations provide sufficient privacy protection to satisfy

[55] The difference of 0.00000000001% between the baseline of 50% and the set threshold of 50.00000000001% determines the adversary's benefit-cost tradeoff. For example, if Eve's goal is to differentiate between the two messages Attack at dawn and Attack at sunset, then she would have to accumulate a number of encrypted messages that is inversely proportional to this difference (i.e., in the order of magnitude of 10^{12} messages). The difference that is deemed to be acceptable may also affect the efficiency of the cryptographic scheme. For instance, it can affect Alice's computational costs. Exactly how much higher the adversary is permitted to go above the baseline probability is often captured as a parameter that can be tuned to different privacy (and computational cost) levels. For instance, if Alice feels that the message she is sending is not especially sensitive, she might decide that it is acceptable for an adversary to win the game with a probability of 50.000001%.

the privacy requirements of FERPA.

2.3 An introduction to legal approaches to privacy

Information privacy laws around the world vary substantially with respect to their scope of coverage and the protections they afford.[56] This discussion focuses on a subset of laws that restrict the release of statistical information about individuals or groups of individuals, whether released as raw data, de-identified data, or statistical summaries. The applicability of such laws typically turns on the definition of terminology such as personal information, personal data, personally identifiable information, or a similar term.[57] If the information to be released falls within a particular law's definition of personal information, then the law typically applies and restricts the disclosure of the information.[58] If it does not meet the particular law's definition of personal information, then the information is often afforded no or minimal protection under that law.[59] In addition, some privacy laws expressly exclude a category of *de-identified information*. Information that has been transformed such that it satisfies the applicable law's de-identification standard can be shared under relaxed conditions. In some cases, de-identified information can even be released publicly without further restriction on use or redistribution.[60]

Definitions of personal information vary considerably across information privacy laws. The inconsistency between definitions and the reliance on ambiguous and narrow interpretations of these definitions are widely cited as weaknesses of the legal framework for privacy protection.[61] It is beyond the scope of this Article to detail all legal approaches to privacy and definitions of personal information currently in place around the world. Instead, we provide an overview of selected approaches in order to illustrate a range of different approaches and definitions, and some of the challenges that have arisen in developing, interpreting, and complying with various regulatory definitions and standards for privacy protection.

[56] For a broad survey and classification of privacy laws across many jurisdictions, see Bert-Jaap Koops, Bryce Clayton Newell, Tjerk Timan, Ivan Škorvánek, Tomislav Chokrevski, & Maša Galič, *A Typology of Privacy*, 38 U. Pa. J. Int'l L. (forthcoming 2016) (surveying constitutional protections and privacy scholarship from nine North American and European countries).

[57] For an overview of various definitions of personal information, see *See* 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 (2011).

[58] *See, e.g.,* Children's Online Privacy Protection Act, 15 U.S.C. § 6502(a)(1) (providing that "[i]t is unlawful for an operator of a website or online service directed to children, or any operator that has actual knowledge that it is collecting personal information from a child, to collect personal information from a child in a manner that violates the regulations . . ."

[59] *See* 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) (finding that many laws "share the same basic assumption—that in the absence of PII, there is no privacy harm. Thus, privacy regulation focuses on the collection, use, and disclosure of PII, and leaves non-PII unregulated.").

[60] *See, e.g.,* HIPAA Privacy Rule, 45 C.F.R. § 164.502(d)(2) (providing that"[h]ealth information that meets the standard and implementation specifications for de-identification . . . is considered not to be individually identifiable health information, i.e., de-identified. The requirements of this subpart do not apply to information that has been de-identified in accordance with the applicable requirements . . .").

[61] *See* 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, 1893 (2011) (concluding that "there is no uniform definition of PII in information privacy law. Moreover, the definitions that do exist are unsatisfactory.").

2.3.1 Selected approaches from the United States

Privacy law in the United States takes a sectoral approach. Many privacy laws are in place, at the federal and state level, and each law is drawn rather narrowly to protect certain types of information in particular contexts.[62] For illustration, consider the Video Privacy Protection Act, which protects the privacy of individuals in records of video sales and rentals, defines *personally identifiable information* (PII) as "information which identifies a person as having requested or obtained specific video materials or services from a video tape service provider."[63] In contrast, the California Confidentiality of Medical Information Act defines *medical information* as "individually identifiable health information about a patient's medical history, mental or physical condition, or treatment" which "must include an element that identifies a person, such as name, address, email address, telephone number, or Social Security number, or that can be combined with other publicly available information to reveal a person's identity."[64] Many laws in the U.S, adopt some variation of a binary definition that depends on whether the information either does or does not "identify a person."[65] There has been substantial uncertainty in determining whether a specific piece of information does in fact identify a person in practice.[66]

In contrast to laws which are defined in scope by the presence of information that "identifies a person," some laws aim to provide a more bright-line standard, by setting forth an exhaustive list of the types of information that the law protects. An example is the Massachusetts data protection regulation, which defines *personal information* with an exhaustive list: "a Massachusetts resident's first name and last name or first initial and last name in combination with any one or more of the following data elements that relate to such resident: (a) Social Security number; (b) driver's license number or state-issued identification card number; or (c) financial account number, or credit or debit card number, with or without any required security code, access code, personal identification number or password, that would permit access to a resident's financial account" but notes that it does not include "information that is lawfully obtained from publicly available information, or from federal, state or local government records lawfully made available to the general public."[67] As another example, the Health Insurance Portability and Accountability Act Privacy Rule[68] provides a safe harbor, by which data can be shared widely once all information from a list of eighteen categories of information have been removed.[69]

Beyond a small subset of such laws that attempt to take a bright-line approach to defining *personally identifiable information*, privacy laws in the U.S. generally employ standards that require case-by-case determinations that rely on some degree of interpretation. These determinations

[62] For a discussion of the evolution and nature of the U.S. sectoral approach to privacy, see Paul M. Schwartz, *Preemption and Privacy*, 118 Yale L.J. 902 (2008).

[63] 18 U.S.C. § 2710(a)(3) (emphasis added).

[64] Cal. Civ. Code §§ 56–56.37.

[65] *See* 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 (2011).

[66] *See, e.g.,* Pineda v. Williams-Sonoma Stores, 246 P.3d 612, 612 (Cal. 2011) (reversing the lower courts and determining that a "cardholder's ZIP code, without more, constitutes personal identification information" within the meaning of the California Song-Beverly Credit Card Act of 1971 "in light of the statutory language, as well as the legislative history and evident purpose of the statute").

[67] 201 C.M.R. 17.02.

[68] 45 C.F.R. Part 160 and Subparts A and E of Part 164.

[69] 45 C.F.R. § 164.514. Note, however, that HIPAA's safe harbor standard creates ambiguity by requiring that the entity releasing the data "not have actual knowledge that the information could be used alone or in combination with other information to identify an individual who is a subject of the information." *Id.*

are complicated by advances in analytical capabilities, the increased availability of data about individuals, and developments in the scientific understanding of privacy. These developments, in combination with limited guidance on interpreting and applying regulatory standards for privacy, have led individual actors who manage personal data to incorporate a wide range of different standards and practices for privacy protection.[70] Recognizing the need for case-specific determinations, the HIPAA Privacy Rule provides an alternative approach that allows data to be shared pursuant to an expert's determination that "generally accepted statistical and scientific principles and methods for rendering information not individually identifiable" have been applied and provision of documentation that "the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information."[71] Practitioners frequently comment on the ambiguity of these standards and the lack of clarity in interpreting definitions such as personally identifiable information.[72]

In a 2012 survey of commentary on the U.S. legal framework for privacy protection, the Federal Trade Commission (FTC) concluded that "the traditional distinction between PII and non-PII has blurred and that it is appropriate to more comprehensively examine data to determine the data's privacy implications."[73] The FTC, which has authority to bring enforcement actions against companies that engage in unfair and deceptive trade practices including practices related to data privacy and security, takes a different approach. The FTC has developed a privacy framework that applies to commercial entities that collect or use consumer data that "can be reasonably linked to a specific consumer, computer, or other device."[74] FTC guidance has set forth a three-part test for determining whether data are "reasonably linkable" under this standard. To demonstrate that data are not reasonably linkable to an individual identity, a company must (1) take reasonable measures to ensure the data are de-identified, (2) publicly commit not to try to re-identify the data, and

[70] *See, e.g.,* Benjamin C.M. Fung, Ke Wang, Rui Chen & Philip S. Yu, *Privacy-Preserving Data Publishing: A Survey of Recent Developments*, 42 ACM Computing Surveys (2010).

[71] 45 C.F.R. § 164.514(b). The Department of Health & Human Services has declined to provide specific instructions for carrying out an expert determination. *See* HHS Office of the Secretary, Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule HHS.gov, http://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/understanding/coveredentities/De-identification/hhs_deid_guidance.pdf (2012) ("No single universal solution addresses all privacy and identifiability issues. Rather, a combination of technical and policy procedures are often applied to the de-identification task. [The Office for Civil Rights (OCR)] does not require a particular process for an expert to use to reach a determination that the risk of identification is very small. However, the Rule does require that the methods and results of the analysis that justify the determination be documented and made available to OCR upon request. The following information is meant to provide covered entities with a general understanding of the de-identification process applied by an expert. It does not provide sufficient detail in statistical or scientific methods to serve as a substitute for working with an expert in de-identification.").

[72] For example, the 2008 rulemaking to update FERPA acknowledged the confusion expressed by commentators regarding the potential applicability of the law's definition of personally identifiable information. *See* 73 Fed. Reg. at 74,830–31 (noting comments from the public that "the standard . . . about whether the information requested is 'linked or linkable' to a specific student was too vague and overly broad and could be logically extended to cover almost any information about a student," "a comprehensive list of indirect identifiers would be helpful," a definition of "the concept of indirect identifiers" is needed, and clarification of "which personally identifiable data elements may be released without consent" should be provided.)

[73] Federal Trade Commission, Protecting Consumer Privacy in an Era of Rapid Change 2 (2012).

[74] *Id.* at vii. Note also that the framework does not apply to "companies that collect only non-sensitive data from fewer than 5,000 consumers a year, provided they do not share the data with third parties." *Id.* at iv.

(3) contractually prohibit downstream recipients from attempting to re-identify the data.[75] The first prong is satisfied when there is "a reasonable level of justified confidence that the data cannot reasonably be used to infer information about, or otherwise be linked to, a particular consumer, computer, or other device."[76] Noting that it will follow the flexible standard that it follows in data security cases,[77] the FTC clarifies that "what qualifies as a reasonable level of justified confidence depends upon the particular circumstances, including the available methods and technologies," as well as "the nature of the data at issue and the purposes for which it will be used."[78] The FTC notes that various technical approaches can be used to satisfy this standard, and that it "encourages companies and researchers to continue innovating in the development and evaluation of new and better approaches to deidentification. FTC staff will continue to monitor and assess the state of the art in de-identification."[79] As such, the FTC's approach is likely to evolve over time in response to the development of new technologies for privacy protection.

2.3.2 Selected approaches from the European Union

Privacy law in the European Union relies on a definition of *personal data* that is broader than corresponding definitions in the United States. The Data Protection Directive, for instance, defines personal data as "any information relating to an identified or identifiable natural person," where an "an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity."[80] Similarly, the new EU General Data Protection Regulation (GDPR) defines *personal data* as "any information relating to a data subject," where a data subject is defined as a person who "can be identified, directly or indirectly, by means reasonably likely to be used."[81] The Directive's provisions do not apply to "data rendered anonymous in such a way that the data subject is no longer identifiable."[82] The GDPR distinguishes between "pseudonymous information," which "could be attributed to a natural person by the use of additional information" and "anonymous information," which is "namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable."[83] The GDPR "does not [] concern the processing of [] anonymous information," while pseudonymous information "should be considered to be information on an identifiable natural person."[84]

Recital 26 of the GDPR aims to clarify this standard, in articulating that "[t]o determine whether a natural person is identifiable, account should be taken of all the means reasonably likely

[75] *Id.* at 21.
[76] *Id.* at 2.
[77] "The [Federal Trade] Commission's approach in data security cases is a flexible one. Where a company has offered assurances to consumers that it has implemented reasonable security measures, the Commission assesses the reasonableness based, among other things, on the sensitivity of the information collected, the measures the company has implemented to protect such information, and whether the company has taken action to address and prevent well-known and easily addressable security vulnerabilities." *Id.* at 21 n.108.
[78] *Id.* at 21.
[79] *Id.* at 21.
[80] Council Directive 95/46/EC, art. 2, 1995 O.J. (L. 281) 31.
[81] Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data, at 41, COM (2012) 11 final (Jan. 25, 2012).
[82] Data Protection Directive, at recital 26.
[83] GDPR, at recital 26.
[84] *Id.*

to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly."[85] In turn, "to ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments."[86]

The Article 29 Working Party which provides advisory guidelines on the Data Protection Directive has clarified the distinction between an *identified* and *identifiable* person as follows: "a natural person can be considered as 'identified' when, within a group of persons, he or she is 'distinguished' from all other members of the group. Accordingly, the natural person is 'identifiable' when, although the person has not been identified yet, it is possible to do it."[87] It is also clear that "the question of whether the individual to whom the information relates is identified or not depends on the circumstances of the case."[88] For instance, "[a] very common family name will not be sufficient to identify someone - i.e. to single someone out - from the whole of a country's population, while it is likely to achieve identification of a pupil in a classroom. Even ancillary information, such as 'the man wearing a black suit' may identify someone out of the passers-by standing at a traffic light."[89]

The Directive also explains that "whereas to determine whether a person is identifiable account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person."[90] The Article 29 Working Party writes that in interpreting this standard, "the cost of conducting identification," "[t]he intended purpose, the way the processing is structured, the advantage expected by the controller, the interests at stake for the individuals, as well as the risk of organisational dysfunctions (e.g. breaches of confidentiality duties) and technical failures should all be taken into account."[91] It also notes that "this test is a dynamic one and should consider the state of the art in technology at the time of the processing and the possibilities for development during the period for which the data will be processed. . . . The system should be able to adapt to these developments as they happen, and to incorporate then the appropriate technical and organisational measures in due course."[92] To provide specific guidance on applying de-identification techniques, the Article 29 Working Party has released an opinion that assesses the strengths and weaknesses of various technical approaches to de-identification.[93] In implementing the Data Protection Directive and applying the standard for determining whether data subjects can be considered "identifiable" or whether the data have been rendered "anonymous," the EU Member States have adopted divergent interpretations.[94] In 2012, the European Council concluded that some Member States, e.g., Denmark, Finland, France, Italy, Spain, and Sweden, "are generally less demanding [than other Member States] with regard to the processing of data that are not immediately identifiable, taking into account the likelihood of the data subject being identified as

[85] *Id.*
[86] *Id.*
[87] Article 29 Data Protection Working Party, Opinion 4/2007 on the Concept of Personal Data, 12251/03/EN 12 (June 20, 2007).
[88] *Id.* at 13.
[89] *Id.*
[90] Data Protection Directive, at recital 26.
[91] Opinion 4/2007 at 15.
[92] *Id.*
[93] Article 29 Data Protection Working Party, Opinion 05/2014 on Anonymisation Techniques (Apr. 10, 2014).
[94] European Council, Evaluation of the Implementation of the Data Protection Directive, Annex 2, at 15 (2012).

well as the nature of the data."[95]

The wide variation among legal definitions of personal information and how they have been interpreted, the gaps created by the narrowness of their scope (particularly within the U.S. framework), ambiguity regarding the context-specific applicability along the boundaries, and the dynamic nature of the definitions and their interpretation in response to technological developments over time, create challenges for demonstrating that the use of a privacy-preserving technology is sufficient to satisfy legal requirements for privacy protection. In the language of the selected laws introduced in this section, this Article aims to overcome these challenges by proposing an approach that can potentially be used to formalize legal definitions such as "information which identifies a person;"[96] "a reasonable level of justified confidence that the data cannot reasonably be used to infer information about, or otherwise be linked to, a particular consumer, computer, or other device;"[97] and "a person who can be identified, directly or indirectly, by means reasonably likely to be used"[98] so that a privacy technology's compliance with the regulations can be rigorously demonstrated. The approach we propose could also potentially be used by regulators and advisory bodies in future assessments regarding whether an emerging privacy technology satisfies regulatory requirements.

2.4 Related research to bridge between legal and computer science approaches to privacy

Numerous research directions have previously sought to address the gaps between legal and technical approaches to privacy protection. Most closely related to this Article is work by Haney et al. modeling the legal requirements for data collected by the U.S. Census Bureau. Their research proposes formal privacy definitions to match the requirements of Title 13 of the U.S. Code, which protects information furnished to the Census Bureau by individuals and establishments.[99] In order to specify a formal model of Title 13's privacy requirements, Haney et al. introduce a technical definition of privacy that is a variant of differential privacy and that is informed by an understanding of the Census Bureau Disclosure Review Board's historical interpretations of Title 13. The privacy definition they present goes beyond protecting the privacy of individuals, which is inherent to the differential privacy definition, to also protect quantitative establishment information from being inferred with a level of accuracy that is "too high." While similar, the analysis in this Article differs from that explored by Haney et al. in significant ways. First, the model of FERPA in this Article aims to capture a large class of interpretations of the law in order to address potential differences in interpretation, whereas the definition presented by Haney et al. adopts a singular understanding of the Census Bureau's Disclosure Review Board's interpretation of Title 13. In addition to being a narrower model of legal requirements, it is one that relies on internal analysis by the Census Bureau Disclosure Review Board, rather than drawing directly from the text of statutes, regulations, and publicly available policy. Whereas Haney et al. describe specific computations that meet the definition they introduce, the technical analysis in this Article shows that a rich class of computations

[95] *Id.*

[96] Video Privacy Protection Act, 18 U.S.C. § 2710(a)(3).

[97] Federal Trade Commission, Protecting Consumer Privacy in an Era of Rapid Change 2 (2012).

[98] Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data, at 41, COM (2012) 11 final (Jan. 25, 2012).

[99] *See* Samuel Haney et al., *Design of Policy-Aware Differentially Private Algorithms*, Proceedings of the VLDB Endowment, Volume 9 (4): 264 (2015).

(i.e., differentially-private computations) meet the privacy requirements of our model. Finally, with this Article, we aim to introduce concepts from the technical literature on privacy and to describe an approach to combining legal and technical analysis in a way that is accessible to a legal audience.

This Article also relates to a line of work that encodes legal requirements for privacy protection using formal logic in order to verify the compliance of technical systems.[100] A prominent line of research uses Nissenbaum's framework of contextual integrity, which models privacy norms in the flow of information between agents in a system.[101] This framework is used to extract norms from legal privacy requirements that can be encoded using the language of formal logic.[102] This approach, which we will refer to as the *formal logic model*, has been used to formalize large portions of laws such as the Gramm-Leach-Bliley Act and the Health Insurance Portability and Accountability Act.[103] The analysis in this Article differs from the formal logic model in several substantive ways. First, this Article relies on a different type formalization, i.e., a privacy game instead of logic formulae. Second, the formal logic model typically avoids directly addressing ambiguity in the law by using under-specified predicates and making the assumption that their exact unambiguous meaning can be defined exogenously.[104] The model does not specify *how to determine* whether these predicates actually hold.[105] In contrast, this Article aims to address ambiguity directly by modeling the requirements of the law very conservatively so as to capture a wide-range of reasonable interpretations. Third, the formal logic model uses predicates, which are functions that evaluate to either true or false (but to no other "intermediate" value). For example, the logic model may use a predicate that evaluates to true when a message contains information about a specific attribute of an individual (e.g., whether Jonas failed the standardized math exam). In contrast,

[100] A survey of logic-based approaches, as well as other approaches, can be found in Paul N. Otto & Annie I. Antn, Addressing Legal Requirements in Requirements Engineering, 15TH IEEE INTERNATIONAL REQUIREMENTS ENGINEERING CONFERENCE 5 (2007).

[101] *See* Helen Nissenbaum: Privacy in Context: Technology, Policy, and the Integrity of Social Life, Palo Alto, CA: Stanford University Press

[102] More specifically, the requirements are encoded in a first-order logic extended with primitive operators for reasoning about time. *See* Adam Barth, Anupam Datta, John C. Mitchell, Helen Nissenbaum: Privacy and Contextual Integrity: Framework and Applications. IEEE Symposium on Security and Privacy 2006: 184-198; Adam Barth, John C. Mitchell, Anupam Datta, Sharada Sundaram: Privacy and Utility in Business Processes. CSF 2007: 279-294; Omar Chowdhury, Andreas Gampe, Jianwei Niu, Jeffery von Ronne, Jared Bennatt, Anupam Datta, Limin Jia, William H. Winsborough: Privacy promises that can be kept: a policy analysis method with application to the HIPAA privacy rule. SACMAT 2013: 3-14; Omar Chowdhury, Limin Jia, Deepak Garg, Anupam Datta: Temporal Mode-Checking for Runtime Monitoring of Privacy Policies. CAV 2014: 131-149.

[103] *See* Henry DeYoung, Deepak Garg, Limin Jia, Dilsun Kirli Kaynar, Anupam Datta: Experiences in the logical specification of the HIPAA and GLBA privacy laws. WPES 2010: 73-82.

[104] For instance, a model might want to restrict the transmission of a message m about an individual q if m contains the attribute t and t is considered non-public information. In doing so, it might use predicates such as $\textbf{contains}(m, q, t)$ and $t \in \textbf{npi}$. Predicates are functions that evaluate to either True or False. The predicates given as example here are drawn from Adam Barth, Anupam Datta, John C. Mitchell, & Helen Nissenbaum, *Privacy and Contextual Integrity: Framework and Applications*, PROCEEDINGS OF THE 2006 IEEE SYMPOSIUM ON SECURITY AND PRIVACY 184 (2006).

[105] Continuing with the examples from Footnote 104, the model does not specify when is it the case that a message m contains an attribute t about individual q and when is it the case that attribute t is non-public information. This is an intentional modeling decision as is stated explicitly in the seminal paper on this approach: "Much of the consternation about GLBA [Gramm-Leach-Bliley Act] revolves around the complex definition of which companies are affiliates and what precisely constitutes non-public personal information. Our formalization of these norms sidesteps these issues by taking the role *affiliate* and the attribute *npi* to be defined exogenously." Adam Barth, Anupam Datta, John C. Mitchell, & Helen Nissenbaum, *Privacy and Contextual Integrity: Framework and Applications*, PROCEEDINGS OF THE 2006 IEEE SYMPOSIUM ON SECURITY AND PRIVACY 184 (2006).

in our model we attempt to also account for cases in which an adversary can use a message to infer partial information about an attribute.[106] Fourth, the formal logic model supports reasoning about inference through explicit rules,[107] which only accounts for types of inferences anticipated by the authors of the model. In contrast, in our model we do not limit the type of inferences that agents can make. Finally, formal logic models might specify that a doctor is allowed to share a specific patient's private medical information with that patient or that researchers may publish aggregate statistics based on private medical information. However, such models do not enable expressing restrictions on the communication of aggregate statistics such as "the average salary of bank managers can be released only if it does not identify a particular individual's salary."[108] The latter is precisely the type of restriction we are hoping to capture through our model; we focus on modeling the degree to which an adversary should be prevented from inferring private information about an individual from an aggregate statistic.

3 Introduction to two privacy concepts: differential privacy and FERPA

In this Article, we demonstrate an approach for bridging between technical and regulatory privacy concepts. In order to illustrate its use, we focus on two specific real-world privacy concepts—one technical (differential privacy) and the other regulatory (FERPA). In this Section, we introduce the two concepts and set forth the definitions that will form the basis of the analysis that will follow in later sections.

We choose to rely on differential privacy in this Article because its rich and developed theory can serve as an initial subject of an examination of how formal notions of privacy can be compared to regulatory standards. In addition, demonstrating that differential privacy is in accordance with privacy laws may be essential for some practical uses of differential privacy.

3.1 Differential privacy

Differential privacy is a nascent privacy concept that has emerged in the theoretical computer science literature, in response to accumulated evidence of the weaknesses of commonly used techniques for privacy protection such as de-identification.[109] Differential privacy, first presented in 2006, is the result of ongoing research to develop a privacy technology that provides robust protection even against unforeseen attacks. Differential privacy by itself is not a single technological

[106] Technically, this inference results in a change in the probability that the adversary correctly guesses the value of the attribute, without necessarily empowering the adversary observer to guess correctly with certainty.

[107] These rules are in the form (T, t), where T is a set of attributes and t is a specific attribute: if an agent knows the value of every attribute in T about some individual q, then the agent also knows the value of attribute t about q. For instance, the rule ({weight, height}, BMI) specifies that if an agent knows the weight and height of some individual, the agent also knows the body mass index of that individual. *See* Adam Barth, Anupam Datta, John C. Mitchell, & Helen Nissenbaum, *Privacy and Contextual Integrity: Framework and Applications*, Proceedings of the 2006 IEEE Symposium on Security and Privacy 184 (2006).

[108] Adam Barth, Anupam Datta, John C. Mitchell, & Helen Nissenbaum, *Privacy and Contextual Integrity: Framework and Applications*, Proceedings of the 2006 IEEE Symposium on Security and Privacy 184 (2006).

[109] This discussion of differential privacy is adapted from Kobbi Nissim, Thomas Steinke, Alexandra Wood, Micah Altman, Aaron Bembenek, Mark Bun, Marco Gaboardi, David O'Brien & Salil Vadhan, *Differential Privacy: A Primer for a Non-technical Audience (Preliminary Version)*, (2017). For the literature on differential privacy, see sources cited *supra* note 5.

solution but a definition (sometimes referred to as a standard) that states a concrete requirement, and technological solutions are said to satisfy differential privacy if they adhere to the definition. As a strong, quantitative notion of privacy, differential privacy is provably resilient to a very large class of potential misuses. Differential privacy therefore represents a solution that moves beyond traditional approaches to privacy, which require modification as new vulnerabilities are discovered.

3.1.1 The privacy definition and its guarantee

In this section, we offer an intuitive view of the privacy guarantee provided by differentially private computations.[110]

Consider a hypothetical individual John who has the opportunity to participate in a study exploring the relationship between socioeconomic status and medical outcomes. All participants in the study must complete a questionnaire encompassing topics such as their location, their health, and their finances. John is aware of re-identification attacks that have been performed on de-identified data. He is concerned that, should he participate in this study, some sensitive information about him, such as his HIV status or annual income, might be revealed by a future analysis based in part on his responses to the questionnaire. If leaked, this personal information could embarrass him, lead to a change in his life insurance premium, or affect the outcome of a future bank loan application.

If an analysis on the data from this study is differentially private, then John is guaranteed that even though his information is used in the analysis, the outcome of the analysis will not disclose anything that is *specific to him*. To understand what this means, consider a thought experiment, which we refer to as *John's privacy-ideal scenario* and illustrate in Figure 3. John's privacy-ideal scenario is one in which his personal information is omitted but the information of all other individuals is provided as input as usual. Because John's information is omitted, the outcome of the computation *cannot* depend on John's specific information.

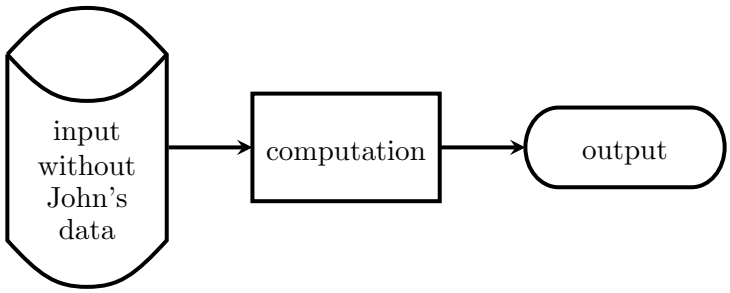

Figure 3: John's privacy-ideal scenario.

Differential privacy aims to provide John with privacy protection in the real-world scenario that approximates his privacy-ideal scenario. Hence, 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 him being included in the computation*. Crucially, this very same guarantee is made not only with respect to John, but also to every other individual contributing his or her information to the analysis.

[110] For the mathematical definition, see the Appendix.

A parameter quantifies and limits the extent of the deviation between the privacy-ideal and real-world scenarios. As shown in Figure 4 below, this parameter is usually denoted by the Greek letter ε (epsilon) and is referred to as the "privacy parameter," or, more accurately, the "privacy loss parameter." The parameter ε measures the effect of each individual's information on the 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 privacy-ideal scenario.[111] Note that in Figure 4 we replaced John with a prototypical individual X to emphasize that the differential privacy guarantee is made simultaneously to *all* individuals in the sample.

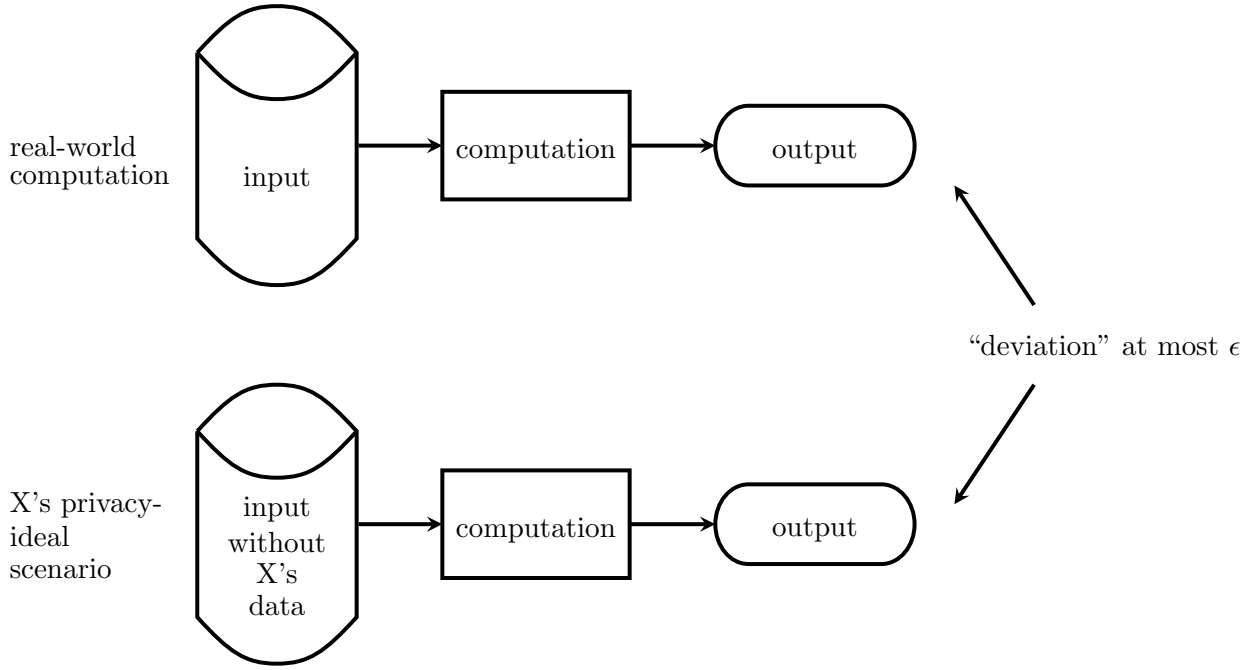

Figure 4: Differential privacy. The maximum deviation between the privacy-ideal scenario and real-world computation should hold simultaneously for each individual X whose information is included in the input.

It is important to note that differential privacy does not guarantee that an observer will not be able to learn anything about John from the outcome of the survey. Consider an observer, Alice, who possesses prior knowledge of some information about John, such as that he regularly consumes a lot of red wine. If the study reports a correlation between drinking red wine and the occurrence of a certain type of liver disease, Alice might conclude that John has a heightened liver disease risk. However, notice that even if information about John is not used in the study, Alice would be able to draw the conclusion that he has a heightened liver disease risk just like other red wine drinkers. In other words, this risk is present in both John's privacy-ideal scenario and his real-world scenario.

John may be adversely affected by the discovery of the results of a differentially private computation (for example, if sales of red wine were made illegal as a result of the discovery). The

[111] For a more detailed discussion of how the privacy parameter ϵ controls risk, see Kobbi Nissim, Thomas Steinke, Alexandra Wood, Micah Altman, Aaron Bembenek, Mark Bun, Marco Gaboardi, David O'Brien & Salil Vadhan, *Differential Privacy: A Primer for a Non-technical Audience (Preliminary Version)*, (2017).

guarantee is that such harm is not due to the presence of John's data; that is, it would occur also in his privacy-ideal scenario.

3.1.2 Differential privacy in the real world

Despite being a relatively new concept, differential privacy has already found use in several real-world applications, and more applications are currently under development. The U.S. Census Bureau makes available an online interface for exploring the commuting patterns of workers across the United States, using confidential data collected through the Longitudinal Employer-Household Dynamics program over the period of 2002–2014.[112] Users of this interface interact with synthetic datasets that have been carefully generated from confidential agency records. The computations used to synthesize the data provide formal privacy guarantees and meet a variant of differential privacy.[113]

Google, Apple, and Uber are also experimenting with differentially private applications. For example, Google's RAPPOR system developed by Google employs differentially private computations to collect information from users of the company's Chrome web browser, in order to gather statistics used to monitor how unwanted software hijacks the browser settings of their users.[114] This application allows analysts at Google to study trends present in the extensive Chrome user base, with strong guarantees that not much can be learned that is specific to any individual user.[115]

The academic community is also in the process of developing practical platforms for performing differentially private analyses. The *Putting Differential Privacy to Work* project at the University of Pennsylvania strives to build a general system that enables the average programmer to employ differentially private computations in a range of applications.[116] As part of the *Privacy Tools for Sharing Research Data* project at Harvard University,[117] differential privacy will be integrated with TwoRavens,[118] a browser-based software interface for exploring and analyzing data that are hosted on the Dataverse repository platform, which is currently used by institutions throughout the world.[119] This will allow researchers to see the results of statistical analyses created by differentially private computations on sensitive datasets, including datasets that cannot otherwise be shared

[112] *See* U.S. Census Bureau, OnTheMap Application for the Longitudinal Employer-Household Dynamics program, http://onthemap.ces.census.gov (last visited Apr. 30, 2016).

[113] *See* Ashwin Machanavajjhala, Daniel Kifer, John Abowd, Johannes Gehrke, & Lars Vilhuber, *Privacy: Theory Meets Practice on the Map*, Proceedings of the IEEE 24th International Conference on Data Engineering 277 (2008).

[114] *See* Úlfar Erlingsson, *Learning Statistics with Privacy, aided by the Flip of a Coin*, Google Research Blog (Oct. 30, 2014), http://googleresearch.blogspot.com/2014/10/learning-statistics-with-privacy-aided.html; Úlfar Erlingsson, Vasyl Pihur & Aleksandra Korolova, *RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response*, Proceedings of the 21st ACM Conference on Computer and Communications Security (2014).

[115] Other examples for using differential privacy (for which, to the best of our 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 Europe Blog (Nov. 18, 2015), http://googlepolicyeurope.blogspot.com/2015/11/tackling-urban-mobility-with-technology.html, and Andy Greenberg, *Apples 'Differential Privacy' Is About Collecting Your Data–But Not Your Data*, Wired (Jun. 13, 2016), http://www.wired.com/2016/06/apples-differential-privacy-collecting-data/.

[116] Putting Differential Privacy to Work project, http://privacy.cis.upenn.edu (last visited Apr. 30, 2016).

[117] Privacy Tools for Sharing Research Data, http://privacytools.seas.harvard.edu (last visited Apr. 30, 2016).

[118] Institute for Quantitative Social Science, TwoRavens, http://datascience.iq.harvard.edu/about-tworavens (last visited Apr. 30, 2016).

[119] Institute for Quantitative Social Science, Dataverse, http://datascience.iq.harvard.edu/about-dataverse (last visited Apr. 30, 2016).

widely due to privacy concerns.

3.2 The Family Educational Rights and Privacy Act of 1974 (FERPA)

FERPA is a U.S. federal law requiring the protection of personal information contained in education records.[120] Education records are defined as records that directly relate to a student,[121] and are maintained by an educational agency or institution that receives funding under a program administered by the U.S. Department of Education.[122] Entities covered by FERPA include elementary and secondary schools, school districts, colleges and universities, state educational agencies, and other institutions providing educational services or directing institutions that do.[123]

FERPA provides parents with certain rights with respect to personal information contained in their child's education records, rights which transfer to an "eligible student" upon turning 18.[124] These rights include the right to inspect, request amendment to, and consent to the disclosure of such information.[125] FERPA distinguishes between two types of personal information contained in education records: *directory information* and *non-directory personally identifiable information*.[126] Generally, a parent or eligible student must provide written consent before an educational agency or institution can disclose non-directory personally identifiable information from a student's education record.[127] Information that a school has designated as directory information can be disclosed without the consent of parents or eligible students, as long as they were provided with notice and an opportunity to opt out of the disclosure of directory information beforehand.

FERPA also permits the disclosure of de-identified information without consent "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."[128] By authorizing the disclosure of de-identified information without restriction, this provision enables the widespread publication and use of statistics on educational programs. According to the Department of Education guidance, this provision is intended to strike "an appropriate balance that facilitates school accountability and educational research while pre-

[120] Family Educational Rights and Privacy Act of 1974, 20 U.S.C. § 1232g; 34 C.F.R. Part 99 (2013).
[121] 20 U.S.C. § 1232g(a)(4).
[122] 20 U.S.C. § 1232g(a)(1)(A).
[123] 34 C.F.R. § 99.1(a)(1-2).
[124] 34 C.F.R. § 99.3.
[125] §§ 99.10, 99.20, 99.30.
[126] This distinction is discussed in depth in Section 3.2.2 below.
[127] *See* § 99.30. The term *disclosure* is defined broadly, meaning "to permit access to or the release, transfer, or other communication of personally identifiable information contained in education records by any means, including oral, written, or electronic means, to any party except the party identified as the party that provided or created the record." 34 C.F.R. § 99.3. Note that FERPA provides a number of exceptions to the written consent requirement, including disclosures to school officials with a legitimate educational interest in the information, §§ 99.31(a)(1), 99.7(a)(3)(iii), disclosures to authorized representatives of the Comptroller General of the U.S., the Attorney General, the Secretary of Education, and State or local educational authorities, §§ 99.31(a)(3), 99.35, and disclosures to organizations conducting studies for, or on behalf of, schools, school districts, or postsecondary institutions, § 99.31(a)(6). However, these exceptions are not a focus of the analysis in this Article because we aim to analyze FERPA's requirements for protecting non-directory personally identifiable information when publishing statistics based on such information. Our analysis of the definition of non-directory personally identifiable information is not affected by exceptions to FERPA allowing the full disclosure of such information to certain parties for specified purposes.
[128] 34 C.F.R. § 99.31(b)(1).

serving the statutory privacy protections in FERPA."[129]

The discussion below introduces the definitions of *directory information*, *personally identifiable information* and *de-identified information* as set forth by the FERPA regulations.[130] These definitions form the basis of our formal model of FERPA's privacy requirements in Section 4. Throughout the analysis, we also refer to guidance from the Department of Education explaining and interpreting the definitions found in the regulations.[131] It is important to note that the Department of Education's interpretations of the regulations may or may not qualify for controlling weight, depending on a number of factors.[132] However, this Article's model of FERPA is based on the regulatory text itself. Interpretations from Department of Education guidance appear in the discussion only for the purposes of demonstrating that the conservative modeling choices capture these interpretations among others falling within a broad class of reasonable interpretations of FERPA.

3.2.1 The applicability of FERPA's requirements to formal privacy models such as differential privacy

FERPA's protections likely apply to the release of statistics, and in particular to releases produced by methods that satisfy formal privacy models such as differential privacy. Therefore, it is important to understand exactly how information privacy laws such as FERPA would govern the use of new technologies based on formal privacy models, in anticipation of the practical implementation of these technologies. In this Section, we point to sources that highlight the ways in which FERPA

[129] 73 Fed. Reg. 74,806, 74,831 (Dec. 9, 2008).

[130] 34 C.F.R. Part 99.

[131] In addition to the regulatory text, this analysis refers to guidance appearing in the preamble to the 2008 final rule to update the FERPA regulations, 73 Fed. Reg. 74,806–55 (Dec. 9, 2008), in which an extended discussion of the definition of *personally identifiable information* was provided in justification of the latest revision to the definition. Note that, while the Department of Education subsequently promulgated rules in 2011, these rules do not amend the definitions of "personally identifiable information" or "de-identified information" set forth under FERPA, nor does the 2011 rulemaking provide guidance on interpreting these concepts. Therefore, these regulations are not pertinent to the analysis in this Article.

[132] As a general rule, courts defer to an agency's interpretation of its own regulations unless the interpretation is "plainly erroneous or inconsistent with the regulation." *See* Auer v. Robbins, 519 U.S. 452, 461 (1997) (quoting Robertson v. Methow Valley Citizens Council, 490 U.S. 332, 359 (1989) (quoting Bowles v. Seminole Rock & Sand Co., 325 U.S. 410, 414 (1945))). However, this rule is limited in several ways. *See* Christopher v. SmithKline Beecham Corp., 132 S. Ct. 2156, 2166 (2012). In particular, an agency's interpretation is not accorded such deference if it "does not reflect the agency's fair and considered judgment on the matter in question," 132 S. Ct. at 2166 (quoting Auer v. Robbins, 519 U.S. at 462), or if the regulation being interpreted merely restates the language of the statute. *See* Gonzales v. Oregon, 546 U.S. 243, 257 (2006). In such cases, the agency's interpretation of its regulations is persuasive rather than controlling. *See* Skidmore v. Swift & Co., 323 U.S. 134, 140 (1944) (holding that the weight of an agency's interpretation is proportional to "the thoroughness evident in its consideration, the validity of its reasoning, its consistency with earlier and later pronouncements, and all those factors which give it power to persuade, if lacking power to control"); Christopher v. SmithKline Beecham Corp., 132 S. Ct. 2156 (2012) (applying the *Skidmore* standard of review to a Department of Labor interpretation of its regulations and finding it to be "quite unpersuasive"). Of the many sources of agency guidance, guidance that appears in the preamble to a final rule, such as the Department of Education's guidance in the preamble to the 2008 final rule interpreting definitions from the amended FERPA regulations that is referenced in this Article, arguably merits greater weight. *See* Kevin M. Stack, *Preambles as Guidance*, 84 Geo. Wash. L. Rev. 1252, 1281 (2016). This is due to various factors, such as the authorship of the preamble by the agency itself, its issuance contemporaneously with the regulations and their statutory justification, and the likelihood that "preamble guidance reflects the kind of deliberate, considered view that would entitle it to the greatest level of deference," as it generally receives a "higher level of vetting internally within the agency and by the executive branch than any other guidance." *See id.*

arguably applies to the release of differentially private statistics. This discussion anticipates and sets the stage for later sections which aim to interpret this language more formally. Although some legal scholars may view the release of aggregate statistics or synthetic data as falling outside the scope of FERPA, we choose to take a more conservative view both because we understand FERPA to apply to the release of aggregate statistics and because we aim to ensure our analysis is valid despite possible ambiguity regarding FERPA's scope of applicability.

Generally, FERPA governs the disclosure of *non-directory personally identifiable information* about students in education records maintained by educational agencies and institutions. Here, *disclosure* is defined broadly, meaning "to permit access to or the release, transfer, or other communication of personally identifiable information contained in education records by any means, including oral, written, or electronic means, to any party except the party identified as the party that provided or created the record."[133] *Personally identifiable information* is also defined broadly to include "information that, alone or in combination, is linked or linkable to a specific student that would allow a reasonable person in the school community, who does not have personal knowledge of the relevant circumstances, to identify the student with reasonable certainty."[134] Because as we discuss above in Section 2.1.3 it has been demonstrated that releases of aggregate statistics can leak some information about individuals, these definitions, which cover any communication of any information linkable to a specific student with reasonable certainty, are arguably written broadly enough to encompass privacy risks associated with statistical releases.

The preamble to the 2008 final rule updating the FERPA regulations also supports the conclusion that FERPA applies to releases of statistical data that adhere to formal privacy models like differential privacy.[135] The Department of Education's rationale in revising FERPA's definition of *personally identifiable information* explicitly addresses privacy issues in releases of statistics. In the preamble, the Department refers to the capability of "re-identifying statistical information or redacted records."[136] By recognizing the privacy risks associated with both "statistical information" and "redacted records," it instructs educational agencies and institutions to consider the privacy risks in both aggregate and individual-level data releases. The preamble provides specific examples to illustrate some of the privacy risks associated with the release of statistical information.[137] The Department of Education notes, for example, that "a school may not release statistics on penalties imposed on students for cheating on a test where the local media have published identifiable information about the only student (or students) who received that penalty; that statistical information or redacted record is now personally identifiable to the student or students because of the local publicity."[138] It also explains how the publication of a series of tables about the same set of students, with the data broken down in different ways, can, in combination, reveal personally identifiable information about individual students.[139] In addition, the guidance notes that educational institutions are prohibited from reporting that 100 percent of students achieved specified performance levels, as a measure to prevent the leakage of personally identifiable information.[140] These references from the preamble to the 2008 final rule are evidence that the Department of

[133] 34 C.F.R. § 99.3.
[134] *Id.* The definition of personally identifiable information is discussed in more detail below in Section 3.2.3.
[135] 73 Fed. Reg. 74,806–55 (Dec. 9, 2008)
[136] 73 Fed. Reg. at 74,832.
[137] *See id.*
[138] *Id.*
[139] *See id.* at 74,835.
[140] *See id.*

Education recognizes privacy risks associated with the release of aggregate statistics.

Educational agencies and institutions share statistics from education records with other agencies, researchers, and the public, for the purposes of research and evaluation. Indeed, the release of certain education statistics is mandated by law,[141] unless such statistics would reveal personally identifiable information as defined by FERPA.[142] For instance, educational agencies and institutions are prohibited from releasing tables containing statistics on groups of individuals falling below some minimum size, where the minimum size varies by state.[143] Each state has established additional procedures for protecting privacy in the release of statistics, including "various forms of suppression, top and bottom coding of values at the ends of a distribution, and limiting the amount of detail reported for the underlying counts."[144] It is important to note, however, that the Department of Education does not consider specific procedures, such as adherence to minimum cell sizes, to be sufficient to meet FERPA's privacy requirements in all cases.[145]

Additionally, the Department of Education's National Center for Education Statistics has released implementation guidance devoted to helping such institutions apply privacy safeguards in accordance with FERPA when releasing aggregate statistics.[146] This guidance aims to clarify the goal of FERPA in the context of aggregate data releases:

> Protecting student privacy means publishing data only in a manner that does not reveal individual students' personally identifiable information, either directly or in combination with other available information. Another way of putting this is that the goal is to publish summary results that do not allow someone to learn information about a specific student.[147]

[141] *See* 20 U.S.C. § 6311(h); 20 U.S.C. § 9607.

[142] 34 C.F.R. § 200.7(b) ("A State may not use disaggregated data for one or more subgroups . . . to report achievement results . . . if the results would reveal personally identifiable information about an individual student . . . [under the requirements of FERPA]").

[143] "Individual states have adopted minimum group size reporting rules, with the minimum number of students ranging from 5 to 30 and a modal category of 10 (used by 39 states in the most recent results available on state websites in late winter of 2010)." National Center for Education Statistics, Statistical Methods for Protecting Personally Identifiable Information in Aggregate Reporting, SLDS Technical Brief: Guidance for Statewide Longitudinal Data Systems (SLDS), https://nces.ed.gov/pubs2011/2011603.pdf (Dec. 2010), at 1.

[144] *Id.*

[145] "[I]t is not possible to prescribe or identify a single method to minimize the risk of disclosing personally identifiable information in redacted records or statistical information that will apply in every circumstance, *including determining whether defining a minimum cell size is an appropriate means to protect the confidentiality of aggregated data and, if so, selection of an appropriate number.* This is because determining whether a particular set of methods for de-identifying data and limiting disclosure risk is adequate cannot be made without examining the underlying data sets, other data that have been released, publicly available directories, and other data that are linked or linkable to the information in question. For these reasons, we are unable to provide examples of rules and policies that necessarily meet the de- identification requirements in 99.31(b). The releasing party is responsible for conducting its own analysis and identifying the best methods to protect the confidentiality of information from education records it chooses to release. . . . With regard to issues with NCLB reporting in particular, determining the minimum cell size to ensure statistical reliability of information is a completely different analysis than that used to determine the appropriate minimum cell size to ensure confidentiality." 73 Fed. Reg. at 74,835 (emphasis added).

[146] *Id.*

[147] *Id.* at 4. Note that one could simply adopt this interpretation and argue that the goal of FERPA's privacy requirements is ''not [to] allow someone to learn information about a specific student," directly implying the differential privacy definition. However, given possible alternatives to the National Center for Education Statistics' interpretation of FERPA, we chose not to adopt this singular interpretation and instead to construct an argument that captures a significantly wider family of potential interpretations of FERPA's requirements, thereby strengthening our argument.

This guidance further "demonstrates how disclosures occur even in summary statistics," discusses how common approaches to privacy may fall short of FERPA's standard for privacy protection, and provides some best practices for applying disclosure limitation techniques in releases of aggregate data.[148] This practical guidance is further evidence that the agency recognizes some leakages of information about individuals from aggregate data releases to be FERPA violations.

Not only does the Department of Education require educational agencies and institutions to address privacy risks in the release of aggregate statistics, but the scientific literature on privacy also suggests that this is a prudent approach. Numerous attacks have demonstrated that it is often possible to link particular individuals to information about them in aggregate data releases.[149] Moreover, the potential leakage of personally identifiable information through releases of aggregate statistics is a concern that is anticipated to evolve and grow over time, as analytical capabilities advance and the availability of large quantities of personal information from various sources increases. It is likely that the approaches identified in current agency guidance will in the future be shown not to provide adequate privacy protection, while also significantly limiting the usefulness of the data released, requiring future updates to the guidance.

For instance, the regulations define *de-identified information* to mean 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."[150] In interpreting this language, the preamble to the 2008 final rule explicitly recognizes that privacy risk from data disclosures is cumulative, and accordingly requires educational agencies and institutions to take into account the accumulated privacy risk from multiple disclosures of information:

> The existing professional literature makes clear that public directories and previously released information, including local publicity and even information that has been de-identified, is sometimes linked or linkable to an otherwise de-identified record or dataset and renders the information personally identifiable. The regulations properly require parties that release information from education records to address these situations.[151]

However, the agency does not provide guidance on addressing cumulative privacy risks from successive disclosures. Rather, it notes that "[i]n the future [it] will provide further information on how to monitor and limit disclosure of personally identifiable information in successive statistical data releases."[152] Indeed, research from the computer science literature demonstrates that it is very difficult to account for the cumulative privacy risk from multiple statistical releases.[153] This example suggests that Department of Education guidance is likely to evolve over time in response to new understandings of privacy risk, particularly with respect to the risk that accumulates from

[148] *Id.* Specifically, best practices identified in the report include "publishing the percentage distribution across categories of outcome measures with no underlying counts or totals; publishing a collapsed percentage distribution across categories of outcome measures with no underlying counts or totals; publishing counts but using complementary suppression at the subgroup level when a small subgroup is suppressed; limiting the amount of detail published for school background information; recoding the ends of percentage distributions; and recoding high and low rates." *Id.*

[149] Privacy risks associated with de-identified and aggregate data are discussed above in Section 2.1.3.

[150] *See* 34 C.F.R. § 99.31(b)(1).

[151] 73 Fed. Reg. at 74,831.

[152] *Id.* at 74,835.

[153] *See* Srivatsava Ranjit Ganta, Shiva Prasad Kasiviswanathan & Adam D. Smith *Composition attacks and auxiliary information in data privacy*, Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 285 (2008).

multiple data releases. It also lends support to the claim that use of differential privacy is sufficient to satisfy FERPA's requirements, as differential privacy provides provable guarantees with respect to the cumulative risk from successive data releases and, to our knowledge, is the only approach to privacy that provides such a guarantee.

3.2.2 The distinction between directory and non-directory information

FERPA distinguishes between *directory information* and *non-directory personally identifiable information*. Directory information is defined as "information contained in an education record of a student that would not generally be considered harmful or an invasion of privacy if disclosed."[154] Each educational agency or institution produces a list of categories of information it designates as directory information. FERPA provides some categories of information for illustration of the types of information an educational agency or institution may designate as directory information:

> Directory information includes, but is not limited to, the student's name; address; telephone listing; electronic mail address; photograph; date and place of birth; major field of study; grade level; enrollment status (e.g., undergraduate or graduate, full-time or part-time); dates of attendance; participation in officially recognized activities and sports; weight and height of members of athletic teams; degrees, honors, and awards received; and the most recent educational agency or institution attended.[155]

Many of these examples of directory information overlap with FERPA's definition of personally identifiable information. However, the law does not require consent for the disclosure of directory information—even directory information constituting personally identifiable information—as long as the relevant educational agency or institution has provided parents and eligible students with public notice of the types of information it has designated as directory information as well as an opportunity to opt out of the disclosure or publication of directory information.[156] An educational agency or institution may also disclose directory information about former students without reissuing the notice and opportunity to opt out.[157]

In contrast, educational agencies and institutions must take steps to protect *non-directory personally identifiable information* from release. This category of information can only be disclosed without consent under certain exceptions, such as the sharing of data for the purposes of developing predictive tests, administering student aid programs, improving instruction, or auditing or evaluating a federal- or state-supported education program.[158] In addition, information from education records that would otherwise be considered non-directory personally identifiable information can be released to the public, without consent, if it has been rendered *de-identified*, meaning "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."[159]

---

[154] 34 C.F.R. § 99.3.
[155] 34 C.F.R. § 99.3.
[156] 34 C.F.R. § 99.37(a).
[157] *Id.*
[158] *See* 34 C.F.R. §§ 99.31(a)(6), (a)(3).
[159] *See* 34 C.F.R. § 99.31(b)(1) ("An educational agency or institution, or a party that has received education records or information from education records under this part, may release the records or information without the consent required . . . after the removal of all personally identifiable information provided that the educational agency or

When releasing data, educational agencies and institutions must assess the privacy-related risks in light of publicly-available information, including directory information.[160] Because the scope of directory information varies from school to school, knowing what information may be available to a potential adversary is uncertain, creating a challenge for an educational agency or institution assessing the privacy risks associated with a planned data release. In Section 4 below, we propose an approach to formally modeling directory information and privacy risks in the release of education data more generally, which addresses the ambiguity created by differences in what may be classified as directory information across different educational agencies and institutions, currently and in the future. Specifically, to address this ambiguity in our model, we make an assumption that a potential privacy attacker is given unrestricted access to information that does not require consent for release, including all potential directory information.

The definition of non-directory personally identifiable information and how it has been interpreted by the Department of Education serves as the basis for the formal model of FERPA's requirements we construct in this Article. Hence, we turn next to the definition of personally identifiable information set forth by the regulations, and how it has been interpreted in agency guidance.

3.2.3 The definition of personally identifiable information

FERPA defines personally identifiable information by way of a non-exhaustive list of categories of information included within the definition. The definition is as follows:

> "Personally Identifiable Information"
>
> The term includes, but is not limited to–
>
> (a) The student's name;
>
> (b) The name of the student's parent or other family members;
>
> (c) The address of the student or student's family;
>
> (d) A personal identifier, such as the student's social security number, student number, or biometric record;
>
> (e) Other indirect identifiers, such as the student's date of birth, place of birth, and mother's maiden name;
>
> (f) Other information that, alone or in combination, is linked or linkable to a specific student that would allow a reasonable person in the school community, who does not have personal knowledge of the relevant circumstances, to identify the student with reasonable certainty; 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.").

[160] The Department of Education observes that "the risk of reidentification may be greater for student data than other information because of the regular publication of student directories, commercial databases, and de-identified but detailed educational reports by States and researchers that can be manipulated with increasing ease by computer technology. . . . [T]he re-identification risk of any given release is cumulative, i.e., directly related to what has previously been released, and this includes both publicly-available directory information, which is personally identifiable, and de-identified data releases." 73 Fed. Reg. at 74,834.

(g) Information requested by a person who the educational agency or institution reasonably believes knows the identity of the student to whom the education record relates.[161]

This analysis focuses in particular on paragraph (f) of the above definition, which forms the basis of the formal model of FERPA's privacy requirements presented in Section 4. Agency guidance from the preamble to the 2008 final rule[162] and other documents[163] provide interpretations of this definition. The definition, by referring to "a reasonable person in the school community"[164] makes use of an objective, reasonableness standard. The reasonable person standard is a common legal standard, referring to a "hypothetical, rational, prudent, average individual."[165] By referring to "a reasonable person in the school community," the Department of Education states that it intended to "provide the maximum privacy protection for students" because "a reasonable person in the school community is also presumed to have at least the knowledge of a reasonable person in the local community, the region or State, the United States, and the world in general."[166] The agency also notes that the standard was not intended to refer to the "technological or scientific skill level of a person who would be capable of re-identifying statistical information or redacted record."[167] Rather, it refers to the knowledge a reasonable person might have "i.e., based on local publicity, communications, and other ordinary conditions."[168] The preamble includes some examples of members of the school community, such as "students, teachers, administrators, parents, coaches, volunteers," and others at the local school.[169]

In 2008, the Department of Education updated the definition of personally identifiable information that appears in the FERPA regulations. Previously, it had included "information that would make the student's identity easily traceable" in lieu of clauses (e)–(g) found in the current definition.[170] The Department of Education explained that it removed the "easily traceable" language from the definition "because it lacked specificity and clarity" and "suggested that a fairly low standard applied in protecting education records, i.e., that information was considered personally identifiable only if it was easy to identify the student."[171]

---

[161] 34 C.F.R. § 99.3.

[162] 73 Fed. Reg. 74,806–55.

[163] *See, e.g.,* National Center for Education Statistics, Statistical Methods for Protecting Personally Identifiable Information in Aggregate Reporting, SLDS Technical Brief: Guidance for Statewide Longitudinal Data Systems (SLDS), https://nces.ed.gov/pubs2011/2011603.pdf (Dec. 2010).

[164] 34 C.F.R. § 99.3.

[165] 73 Fed. Reg. at 74,832.

[166] *Id.*

[167] *Id.* at 74,831.

[168] *Id.* at 74,832.

[169] *Id.*

[170] The previous definition of personally identifiable information, promulgated in 1988, appeared as follows:

> "Personally identifiable information" includes, but is not limited to—
> (a) The student's name;
> (b) The name of the student's parent or other family member;
> (c) The address of the student or student's family;
> (d) A personal identifier, such as the student's social security number or student number;
> (e) A list of personal characteristics that would make the student's identity easily traceable; or
> (f) Other information that would make the student's identity easily traceable.
>
> 53 Fed. Reg. 11943 (Apr. 11, 1988).

[171] 73 Fed. Reg. at 74,831.

The preamble also explains the adoption of an objective, reasonableness standard, in light of the revision to the definition of personally identifiable information. FERPA's reasonableness standard is not to be interpreted as subjective, or based on the motives or capabilities of a potential attacker.

> The "reasonableness" standards in paragraphs (f) and (g) of the new definition, which replace the "easily traceable" standard, do not require the exercise of subjective judgment or inquiries into a requester's motives. Both provisions require the disclosing party to use legally recognized, objective standards by referring to identification not in the mind of the disclosing party or requester but by a reasonable person and with reasonable certainty, and by requiring the disclosing party to withhold information when it reasonably believes certain facts to be present. These are not subjective standards, and these changes will not diminish the privacy protections in FERPA.[172]

In providing guidance on interpreting the definitions of personally identifiable information and de-identified information, the Department of Education acknowledges that FERPA's privacy standard requires case-by-case determinations. It notes that the removal of "nominal or direct identifiers," such as names and Social Security numbers, "does not necessarily avoid the release of personally identifiable information."[173] Furthermore, the removal of other information such as address, date and place of birth, race, ethnicity, and gender, may not be sufficient to prevent one from "indirectly identify[ing] someone depending on the combination of factors and level of detail released."[174] However, the preamble declines "to list all the possible indirect identifiers and ways in which information might indirectly identify a student" because "[i]t is not possible" and "[i]n order to provide maximum flexibility to educational agencies and institutions."[175] In this way, FERPA's privacy requirements "unlike the HIPAA Privacy Rule, [] do not attempt to provide a 'safe harbor' by listing all the information that may be removed in order to satisfy the de-identification requirements []." The preamble also emphasizes that de-identified information that is released could be linked or linkable to "public directories and previously released information, including local publicity and even information that has been deidentified," rendering it personally identifiable.[176] The agency concludes that "[t]he regulations properly require parties that release information from education records to address these situations."[177] Ultimately, "determining whether a particular set of methods for de-identifying data and limiting disclosure risk is adequate cannot be made without examining the underlying data sets, other data that have been released, publicly available directories, and other data that are linked or linkable to the information in question. For these reasons, we are unable to provide examples of rules and policies that necessarily meet the de-identification requirements []. The releasing party is responsible for conducting its own analysis and identifying the best methods to protect the confidentiality of information from education records it chooses to release."[178]

Examples provided in the preamble contribute to a lack of clarity around applying FERPA's privacy requirements. For instance, it is not clear how an educational agency or institution should differentiate between special knowledge and information known to a reasonable person in the school

[172] *Id.*
[173] *Id.* at 74,831.
[174] *Id.*
[175] *Id.* at 74,833.
[176] *Id.*
[177] *Id.*
[178] *Id.* 74, 835.

community. The agency provides the following example to illustrate how the language "personal knowledge of the relevant circumstances," found in paragraph (f) of the definition of personally identifiable information, is to be interpreted:

> [I]f it is generally known in the school community that a particular student is HIV-positive, or that there is an HIV-positive student in the school, then the school could not reveal that the only HIV-positive student in the school was suspended. However, if it is not generally known or obvious that there is an HIV-positive student in school, then the same information could be released, even though someone with special knowledge of the student's status as HIV-positive would be able to identify the student and learn that he or she had been suspended.[179]

This example seems counterintuitive because it does not address whether members of the school community might know, or might in the future learn, that a particular student was suspended. Possession of such knowledge would enable one to learn that this student is HIV-positive. Enabling this type of disclosure through a release of information—highly sensitive information such as a student's HIV status, no less—is likely not what the agency intended. However, it is not clear what the agency did in fact intend to convey with this example. In another example, the agency notes that "if teachers and other individuals in the school community generally would not be able to identify a specific student based on the student's initials, nickname, or personal characteristics contained in the record, then the information is not considered personally identifiable and may be released without consent."[180] This seems to imply a weak privacy standard, as a student's "initials, nickname, or personal characteristics" are likely to be uniquely identifying in many cases, regardless of whether such characteristics are considered to be generally known within the community.

Ambiguity in interpreting this definition is also reflected in a discrepancy between the preamble and an interpretation of the regulation developed by the Privacy Technical Assistance Center (PTAC), a Department of Education contractor that develops guidance on complying with FERPA's requirements.[181] In guidance on interpreting the standard used to evaluate disclosure risk when releasing information from education records, PTAC advises that "[s]chool officials, including teachers, administrators, coaches, and volunteers, are not considered in making the reasonable person determination since they are presumed to have inside knowledge of the relevant circumstances and of the identity of the students."[182] This interpretation from PTAC appears to be substantially weaker than the regulatory text and directly contradicts the language of the 2008 final rule, which interprets the regulations to protect non-directory personally identifiable information from disclosure if it can be identified by this same constituency of the school community. Such individuals would seem to be "reasonable person[s] in the school community," based on the plain meaning of this language. In the preamble, the Department of Education provides the following example: "[I]t might be well known among students, teachers, administrators, parents, coaches, volunteers, or others at the local high school that a student was caught bringing a gun to class last month but generally unknown in the town where the school is located. In these circumstances, a school district may not disclose that a high school student was suspended for bringing a gun to class last

[179] 73 Fed. Reg. 74,832.
[180] *Id.* at 74,831.
[181] *See* Privacy Technical Assistance Center, About PTAC, http://ptac.ed.gov/about.
[182] Privacy Technical Assistance Center, Frequently Asked Questions—Disclosure Avoidance (2015), http://ptac.ed.gov/sites/default/files/FAQ_Disclosure_Avoidance.pdf (last updated May 2013).

month, even though a reasonable person in the community where the school is located would not be able to identify the student, because a reasonable person in the high school would be able to identify the student."[183] As discussed above, a court would likely give more weight to the agency's interpretation in the preamble to the final rule than to an agency contractor's subsequent interpretation. This example illustrates the potential for alternative interpretations of the regulatory requirements.

Numerous commentators have likewise expressed uncertainty regarding interpretations of the privacy requirements of FERPA. For example, the preamble to the 2008 final rule refers to many public comments seeking clarification on the de-identification standard. Comments refer to the standard as being "too vague and overly broad," as the definition of personally identifiable information "could be logically extended to cover almost any information about a student."[184] Other commenters question whether the standard provides privacy protection as strong as the agency intends, in light of concerns about the difficulty of de-identifying data effectively.[185] One commenter noted the ambiguity of the standard, but viewed it in a positive light, by arguing that "ambiguity in the terms 'reasonable person' and 'reasonable certainty' was necessary so that organizations can develop their own standards for addressing the problem of ensuring that information that is released is not personally identifiable."[186]

In Section 4, we aim to overcome ambiguities in the regulatory requirements by modeling them formally, based on conservative, "worst-case" assumptions. We believe this approach can be used to demonstrate that a privacy technology satisfies any reasonable interpretation (or at least a large family of reasonable interpretations) of FERPA's requirements. To construct this formal model of FERPA, we look to the regulatory definitions and, to a lesser extent, agency interpretations of these definitions. In particular, our model of the FERPA standard aims to honor the Department of Education's intent to "provide the maximum privacy protection for students,"[187] by providing protection against a strong adversary. Such an adversary is not constrained by a limited "technological or scientific skill level,"[188] possesses knowledge about students that "a reasonable person in the school community"[189] might have, potentially has the capacity to learn something about an individual in the data even if the individual's identity is not "easily traceable,"[190] and has motives

[183] 73 Fed. Reg. 74,832.

[184] *See* 73 Fed. Reg. at 74,830–34 (including comments such as "the standard . . . about whether the information requested is 'linked or linkable' to a specific student was too vague and overly broad and could be logically extended to cover almost any information about a student," "'relevant circumstances' in paragraph (f) is vague," "a comprehensive list of indirect identifiers would be helpful," a definition of "the concept of indirect identifiers" is needed, clarification of "which personally identifiable data elements may be released without consent" should be provided, "the regulations should provide objective standards for the de-identification of education records," and "examples to help districts determine whether a nontargeted request will reveal personally identifiable information" are needed).

[185] *See id.* at 74,833–834 (referring to comments noting that "complete de-identification of systematic, longitudinal data on every student may not be possible," that "many institutions and individuals have the ability to re-identify seemingly deidentified data and that it is generally much easier to do than most people realize because 87 percent of Americans can be identified uniquely from their date of birth, five-digit ZIP code, and gender," and that "re-identification is a much greater risk for student data than other kinds of information because FERPA allows for the regular publication of student directories that contain a wealth of personal information, including address and date of birth, that can be used with existing tools and emerging technology to re-identify statistical data, even by non-experts").

[186] *Id.* at 74,830.

[187] 73 Fed. Reg. at 74,832.

[188] *Id.* at 74,831.

[189] 34 C.F.R. § 99.3.

[190] 73 Fed. Reg. at 74,831.

that are unknown.[191] The model we outline below does not require a determination of the subjective judgment or inquiries into an attacker's motives and declines to presuppose the attacker's goal.[192] In this way, the model is able to consider attackers with different goals, different pieces of outside knowledge about students, and different ways of using the information, consistent with the flexible, case-by-case approach taken by the FERPA regulations.[193]

3.3 Gaps between FERPA and differential privacy

The emergence of formal privacy models such as differential privacy represents a shift in the conceptualization of data privacy risks and ways to mitigate such risks. The FERPA regulations and guidance from the Department of Education were drafted, for the most part, prior to the development and practical implementation of formal privacy models. They are largely based on traditional approaches to privacy through, for example, their emphasis on the importance of removing personally identifiable information from data prior to release. The regulatory approach therefore differs from the approach relied upon by formal privacy models, creating challenges for translating between the two notions. To illustrate the gap between these two privacy concepts, we outline a number of the key differences below. Note, however, that this discussion is limited, in that it is an illustration of the challenges of applying interpretations of FERPA to implementations of differential privacy. It is not intended as a more general assessment or critique of the level of privacy protection provided by FERPA.

Overall scope of privacy protection. FERPA does not apply across the board to protect all types of data in all settings. Instead, it is designed to protect certain types of information in specific contexts. For instance, FERPA applies only to educational agencies and institutions and protects only certain types of information from education records known as non-directory personally identifiable information. In addition, it primarily appears to address releases of information that could be used in record linkage attacks, or the linkage of a named individual to a record in a release of data, that leverage publicly available data. Differential privacy, in contrast, is designed to be broadly applicable, providing formal bounds on the leakage of *any* information about an individual, not just an individual's identity or non-directory personally identifiable information. Given its comparatively narrow scope, FERPA's applicability to a more general definition of privacy like differential privacy is arguably unclear.

Range of attacks. While guidance on interpreting the law expresses the Department of Education's intent to provide strong privacy protection that addresses a wide range of privacy risks, in effect the regulations seem to address a narrower category of attacks. By permitting the release of de-identified information, or information from which certain pieces of information have been removed, these provisions seem primarily aimed at addressing record linkage attacks that could, using certain data known to be available, enable the linkage of a named individual with a record in a released set of data. In contrast, differential privacy is a quantitative *guarantee* of privacy that is provably resilient to a very large class of potential data misuses. Its guarantee holds no matter what computational techniques or resources the privacy adversary brings to bear. In this way, differential privacy provides protection against inference attacks and attacks unforeseen at the time the privacy-preserving technique is applied. Because the FERPA regulations and implementation

[191] *See id.*
[192] *See id.*
[193] *See id.* at 74,833-35.

guidance focus on certain types of known privacy attacks, it is not clear how the regulations apply to formal privacy models which provide more general protection, including protection against inference attacks and attacks currently unknown.

Scope of private information. FERPA's privacy requirements focus on protecting personally identifiable information. They draw a sharp binary distinction between non-directory personally identifiable information and de-identified information, protecting the former but deeming the latter to fall outside of the scope of the regulations altogether. The regulatory definition of personally identifiable information provides a non-exhaustive list of the categories of information that should be protected from release. Accordingly, a common practice is for educational agencies and institutions is to withhold or redact certain pieces of information, such as names, Social Security numbers, and addresses when disclosing information from education records. The literature recognizes, however, that privacy risks are not limited to certain categories of information; indeed, information not typically used for identification purposes can often be used to identify individuals in de-identified data.[194] For example, a small number of data points about an individual's characteristics, behavior, or relationships can be sufficient to identify an individual.[195] Moreover, a wide range of types of inferences of personal information about an individual, not just an individual's identity, can be made based on a release of de-identified information. Differential privacy takes this broader conception of private information into account by putting formal bounds on any leakage of any information about an individual from a system. How FERPA's binary conceptions of non-directory personally identifiable information and de-identification apply to a formal privacy model, which bounds the incremental leakage of any information specific to an individual, is uncertain.

Form of data release. By relying on terminology such as personally identifiable information and de-identification, the FERPA regulations seems to be written with microdata, or individual-level data, as their primary use case. To a lesser extent, guidance on interpreting FERPA refers to protecting information in releases of statistical tables, and it is limited to the specification of minimum cell sizes and related approaches from the traditional statistical disclosure limitation literature.[196] In addition, by referring explicitly to de-identification and permitting the disclosure of information from which categories of non-directory personally identifiable information have been removed, FERPA appears to endorse heuristic de-identification techniques, such as redaction of pieces of information deemed to be direct or indirect identifiers. These references to de-identification, approaches to risk in microdata releases, and traditional disclosure limitation techniques are difficult to generalize to other types of techniques. For techniques that rely on formal privacy models like differential privacy, which address risk in non-microdata releases, it is not clear how FERPA's privacy requirements should be applied.

Scope of guidance. The FERPA regulations and implementation guidance focus on practices for protecting information in cases where the risks to individual privacy are clear. Consider, for example, FERPA's definition of personally identifiable information, which includes a non-exhaustive list

[194] Narayanan and Shmatikov make an even bolder statement: "[a]ny information that distinguishes one person from another can be used for re-identifying anonymous data." *See* Arvind Narayanan & Vitaly Shmatikov, *Myths and Fallacies of "Personally Identifiable Information,"* 53 Communications of the ACM 24, 26 (2010).

[195] *See* Yves-Alexandre de Montjoye et al., *Unique in the Shopping Mall: On the Reidentifiability of Credit Card Metadata*, 347 Science 536 (2015); Yves-Alexandre de Montjoye et al., *Unique in the Crowd: The Privacy Bounds of Human Mobility*, 3 Nature Sci. Reps. 1376 (2013).

[196] *See* 73 Fed. Reg. at 74,835 (directing educational agencies and institutions to consult the Federal Committee on Statistical Methodologys Statistical Policy Working Paper 22 for guidance on applying methods for protecting information in a data release).

of identifiers such as names, addresses, Social Security numbers, dates of birth, places of birth, and mother's maiden names.[197] When de-identifying data prior to release, an educational agency or institution must take steps to suppress fields containing these categories of information. It is more difficult to determine how to protect other information also falling within this definition, such as "information that, alone or in combination, is linked or linkable to a specific student that would allow a reasonable person in the school community, who does not have personal knowledge of the relevant circumstances, to identify the student with reasonable certainty."[198] Indeed, the Department of Education acknowledges this, stating that "[i]t is not possible, however, to list all the possible indirect identifiers and ways in which information might indirectly identify a student."[199] Educational agencies and institutions are provided little guidance on the steps necessary for protecting privacy in accordance with FERPA beyond redacting certain categories of obvious identifiers. The guidance is particularly unclear regarding how to interpret this language with respect to statistical computations, which do not contain direct or indirect identifiers. Where guidance refers to aggregate data, it flags the disclosure risks associated with the release of statistics derived from as few as one or two individuals.[200] Making a determination whether the steps that have been taken to protect privacy are sufficient, once these clear categories of concern have been addressed, is left as a flexible, case-specific analysis by design.[201] Because FERPA does not set forth a general privacy goal—one that can be demonstrably satisfied with certainty for any statistical computation—it is difficult to determine when a privacy-preserving technology provides protection that is sufficient to satisfy the regulatory requirements.

On first impression, it seems likely that the conceptual differences between FERPA's privacy requirements and differential privacy are too great to be able to make an argument that differentially private tools can be used to satisfy FERPA. However, in the sections that follow, we detail such an argument. Moreover, other regulations such as the HIPAA Privacy Rule share some of the characteristics of FERPA identified in this Section, suggesting that parts of the analysis presented in this article could be applied to make similar arguments with respect to other information privacy laws.

---

[197] 34 C.F.R. § 99.3.

[198] *Id.*

[199] 73 Fed. Reg. at 74,833.

[200] *Id.* at 74,834.

[201] *Id.* at 74,835 ("In response to requests for guidance on what specific steps and methods should be used to de-identify information [], it is not possible to prescribe or identify a single method to minimize the risk of disclosing personally identifiable information in redacted records or statistical information that will apply in every circumstance, including determining whether defining a minimum cell size is an appropriate means to protect the confidentiality of aggregated data and, if so, selection of an appropriate number. This is because determining whether a particular set of methods for de-identifying data and limiting disclosure risk is adequate cannot be made without examining the underlying data sets, other data that have been released, publicly available directories, and other data that are linked or linkable to the information in question. For these reasons, we are unable to provide examples of rules and policies that necessarily meet the de- identification requirements in 99.31(b). The releasing party is responsible for conducting its own analysis and identifying the best methods to protect the confidentiality of information from education records it chooses to release. We recommend that State educational authorities, educational agencies and institutions, and other parties refer to the examples and methods described in the NPRM [] and refer to the Federal Committee on Statistical Methodologys Statistical Policy Working Paper 22 [].").

3.4 Value in bridging these gaps

Legal and computer science concepts of privacy are evolving side by side, and it is becoming increasingly important to understand how they can work together to provide strong privacy protection in practice. The evolution of concepts and understandings of privacy in the the field of computer science can benefit from an understanding of normative legal and ethical privacy desiderata. Similarly, legal privacy scholarship can be informed and influenced by concepts originating in computer science thinking and by the understanding of what privacy desiderata can be technically defined and achieved. We argue, however, that to promote a mutual influence between the fields it is necessary to overcome the substantial gaps between the two approaches. Furthermore, in light of the vast bodies of theory and scholarship underlying the concepts, we believe an approach that is rigorous from both a legal and a technical standpoint is required.

Bridging the gap between technical and regulatory approaches to privacy can help support efforts to bring formal privacy models such as differential privacy to practice. Uncertainty about compliance with strict regulatory requirements for privacy protection may act as barriers to adoption and use of emerging techniques for analyzing sensitive information in the real world. If data holders, owners, or custodians can be assured that the use of formal privacy models will satisfy their legal obligations, they will be more likely to begin using such tools to make new data sources available for research and commercial use.

At the same time, this interdisciplinary approach is also important for the future of robust privacy regulation. Because information privacy is in large part a highly technical issue, it will be critical for policymakers to understand the privacy technologies being developed and their guarantees. This understanding is needed in order to approve and guide the use of formal privacy models as a means of satisfying regulatory requirements. In addition, a clear understanding of the principles underlying formal approaches to privacy protection and the ways in which they differ from the concepts underlying existing regulatory definitions, and technical approaches outlined in current agency guidance, help illustrate the weaknesses in the regulatory framework and point to a new path forward. Taken together, these insights can be used to help pave the way for the adoption of robust privacy practices in the future.

4 Extracting a formal privacy definition from FERPA

The goal of this section is to extract a mathematical model of FERPA's privacy requirements. We can use such a model to formally prove that privacy technologies that adhere to a model such as differential privacy meet the requirements of the law. We provide an example of such a proof later in this Article.

The approach we take is inspired by the game-based privacy definitions used in the field of computer science. As discussed in Section 2.2 above, a game-based approach defines privacy via a hypothetical game in which an adversary attempts to learn private information based on the output of a computation performed on private data. If it can be shown that the adversary cannot win the game "too much," the computation is considered to protect privacy. Recall, for example, the privacy game from Section 2.2. In this scenario, Alice seeks to send an encrypted message to Bob with confidence that an eavesdropper cannot learn much about the content of their communication by looking at the encrypted text. Alice is able to formalize her privacy desiderata as a privacy game, and she can use encryption algorithms with confidence as long as they have been mathematically proven to meet her definition of privacy.

Analogously, in order to demonstrate that a given privacy technology meets the privacy requirements of FERPA, the first step is to define a game that faithfully encompasses the privacy desiderata that were envisioned by the Department of Education when drafting the regulations. To do so, we must carefully define the capabilities of the adversary and the mechanics of the game in ways that capture the privacy threats that FERPA was designed to protect against. Because FERPA is written in the language of regulation and not as a formal mathematical definition, it is open to different, context-dependent interpretations by design. To deal with the inherent ambiguity in the language, our desideratum is to design a game that conservatively accounts for (ideally) any reasonable interpretation of the regulatory text, or, at least, accounts for a very large class of such interpretations. As we will describe in detail below, this requires us to design games that give the adversary what might be considered to be an unrealistic advantage. However, if we can prove that a system provides privacy protection given extremely conservative assumptions, we will have also proven that it provides privacy protection also in the more realistic scenarios.

We begin with a simple, informal view of a privacy game based on FERPA's requirements. We can imagine a game in which a school classifies the student information it maintains into two distinct categories as defined by FERPA: *directory information* (which can be made publicly available in accordance with FERPA, and therefore in our modeling we assume it is available to the attacker) and *non-directory personally identifiable information* (which FERPA protects from disclosure, so we do not assume it is available to the attacker). In the game, a statistical computation is performed by the game mechanics over this information and the result is shown to the adversary. The adversary wins the game if she can successfully guess a sensitive attribute of a student, i.e., link a named student to non-directory personally identifiable information about that student. Figure 5 represents this game visually.[202]

To make this game framework more concrete, consider the following hypothetical scenario. The National Center for Education Statistics has decided to publicly release a dataset containing information from education records obtained from schools across the United States. To protect the students' privacy, statisticians have de-identified the dataset using a technique such as an k-anonymization algorithm. After the de-identified dataset has been published online, a data broker attempts to glean non-directory personally identifiable information about students from the k-anonymized data with the goal of selling the re-identified information to marketers. If the data broker is able to successfully extract a student's non-directory personally identifiable information from the release (perhaps by using information it has obtained from other databases), then the privacy of that student has been violated and the data broker's attack has been successful. In the language of a privacy game, the data broker is playing the role of the *adversary* and the k-anonymization algorithm is the *computation* that is intended to provide privacy protection. The data broker *wins the game* by performing a sequence of reasoning, at the end of which the data

[202] The description here omits important details that will be introduced in the following sections. These include the process of creating the directory and non-directory personally identifiable information available to the attacker and a discussion of whether the attacker chooses the student to attack before or after seeing the computation result. Also note that the mathematical formalization of a game abstracts the adversary as an arbitrary computation. This is a reasonable abstraction considering that to optimize its success in winning the game, an adversary needs to perform a sequence of reasonings based on all the information it has gathered on students (including, in particular, the adversary's a priori knowledge, the directory information, and the computation result). We emphasize, however, that by modeling the adversary as an *arbitrary* computation we make no assumptions regarding this sequence of reasonings, and in particular, it does not have to adhere to any known attack strategy. This abstraction is necessary for making our mathematical arguments concrete and precise. Furthermore, the abstraction becomes useful in the proof of privacy, where a hypothetical computation is created that uses the adversary computation as a sub-procedure.

broker successfully guesses a student's non-directory personally identifiable information.

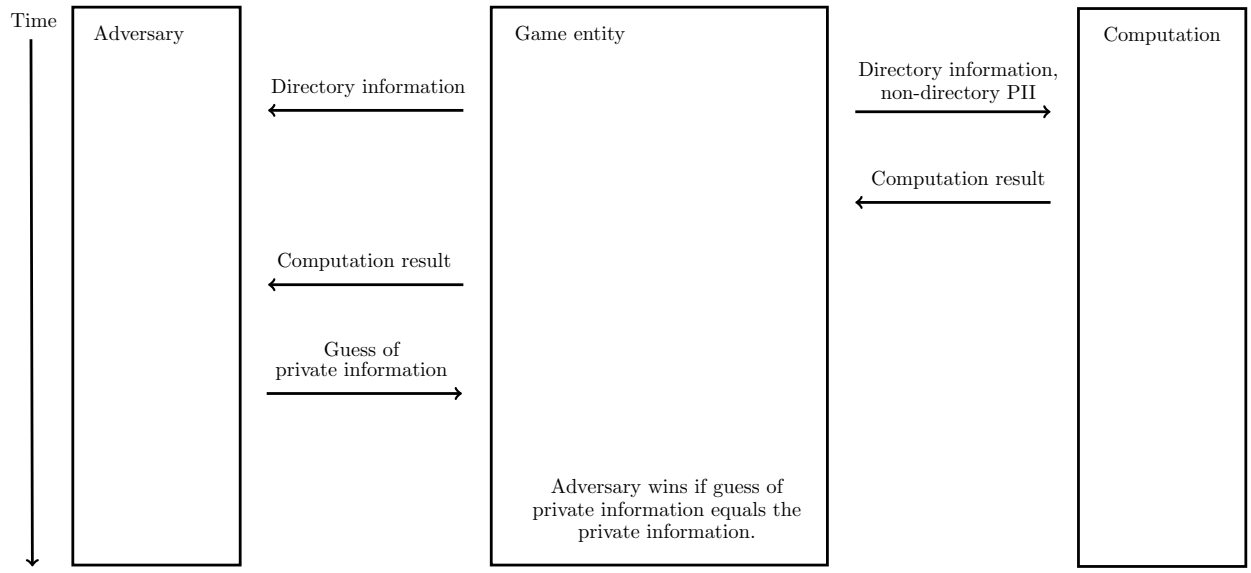

Figure 5: Simplified FERPA game. Note that time progresses from top to bottom, illustrating the sequencing of steps in the game.

Note, however, that a privacy breach does not necessarily occur any time an adversary wins the privacy game. For instance, consider a game in which gender is treated as a private attribute of a student. If challenged to guess the gender of an arbitrary student, an adversary can be expected to win the game with a probability of 0.5, even if the adversary's guess is based on a random coin toss. The adversary will have this probability of success, even if she does not know anything about the student or have access to any statistic computed on that student's private information. If the student's name were publicly available, the adversary might have an even greater chance of success. For example, if the student's given name was "Susan," the adversary could correctly guess the student's gender with probability close to 1. While these examples involving guesses of a student's gender might seem contrived, it is important to acknowledge that, even in richer domains, the adversary always has some likelihood of correctly guessing private information about a student and thus winning the privacy game, even without seeing the results of computations performed on that student's private information. A privacy breach has not occurred in such scenarios because the adversary's successful guess was not informed by a leak of private information. The adversary has won by chance instead. In Section 4.7.2, we discuss how to account for the adversary's baseline probability of success, that is, the adversary's probability of success prior to seeing the results of the computation.

The subsections that follow present an approach to formalizing a game-based privacy definition for FERPA. Justifications for the assumptions made in developing this definition, and an outline of the game-based formalization itself, are also provided. The resulting model attempts to capture a broad scope of potential interpretations of the privacy risks that FERPA is intended to protect against. In this way, a computation that can be proven to satisfy the resulting model can be used to analyze private data or release statistics with a high degree of confidence that the computation adheres to the privacy requirements of FERPA.

4.1 A conservative approach to modeling

Although statutory and regulatory definitions are generally more precise than language used in everyday interactions, they are nevertheless ambiguous. On the one hand, this ambiguity is an advantageous feature, since it builds flexibility into legal standards and leaves room for interpretation, value judgments, and adaptability to new scenarios as practices and social norms inevitably evolve over time. But on the other hand, technological solutions that rely on formal mathematical models require the specification of exact, unambiguous definitions. This presents a challenge for the practical implementation of privacy technologies. How can a legal standard such as the privacy protection required by FERPA be translated into precise mathematical concepts that a technological tool can be evaluated against, or designed to satisfy?

Consider, for instance, a school's disclosure of the Social Security numbers of its students. FERPA explicitly classifies a student's Social Security number as non-directory personally identifiable information and bars it from release.[203] Given this prohibiton, it would clearly be unacceptable to publicly release all nine digits of a student's Social Security number. It would also clearly be acceptable to release zero digits of the number, as such a release would not leak any information about the Social Security number. But would it be acceptable to release three, four, or five digits of a Social Security number? In other words, at what point between disclosing zero and nine digits, is the threshold between an acceptable data release and a prohibited data release?[204]

One way to approach this problem would be to analyze the text of a statute or regulation and decide on a reasonable interpretation as applied to a given privacy technology. However, an interpretation that seems reasonable to one person might not seem reasonable to another. Indeed, judges, lawyers, and legal scholars frequently disagree when interpreting and applying statutory or regulatory language. There is a concern that, if our model of FERPA's privacy requirements were based on a particular interpretation of the law, this interpretation could be disputed by other legal experts. In addition, future applications of the law may lead to new interpretations that challenge the interpretation we have adopted for our model.

To overcome these issues, our model aims to err on the side of a very conservative interpretation of the regulation's privacy requirements. That is, wherever there is a choice between different interpretations of FERPA's requirements, the more restrictive interpretation is selected. For instance, it is not clear from the regulatory requirements how strong of an adversary they must withstand.[205] Therefore, we assume in our model that the adversary is very strong, i.e., well-resourced and capa-

[203] 34 C.F.R. § 99.3.

[204] Note that while this example may appear to be a purely hypothetical concern, regulators do in fact grapple with this type of problem. For instance, guidance from the Department of Health and Human Services on de-identifying data in accordance with the HIPAA Privacy Rule's safe harbor standard for de-identification states that in general "parts or derivatives" of identifiers cannot be released. However, the standard permits the first three digits of a ZIP code to be safely released without violating patients' privacy, as long as the populations of all ZIP codes that begin with those three digits sum to over 20,000 individuals. *See* HHS Office of the Secretary, Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule HHS.gov, https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/understanding/coveredentities/De-identification/hhs_deid_guidance.pdf (2012).

[205] Note, for example, that the Department of Education interprets the definition of personally identifiable information not to clearly limit the technological or skill level of a potential adversary. *See* 73 Fed. Reg. at 74,831 ("The standard proposed in paragraph (f) [of the definition of personally identifiable information] regarding the knowledge of a reasonable person in the school or its community was not intended to describe the technological or scientific skill level of a person who would be capable of re-identifying statistical information or redacted records.").

ble of carrying out a sophisticated privacy attack. This approach effectively strengthens the claim we are making that a given technical approach to privacy protection satisfies a particular legal requirement of privacy protection. If a system can be proved to be secure in strongly antagonistic circumstances, including unrealistically antagonistic circumstances, it is certainly secure given assumptions that are more realistic.

In Figure 6, we provide a visual representation of this desiderata for our model. The figure recognizes that there could be many possible interpretations of the privacy requirements of FERPA. Each possible interpretation can be understood as specifying a set of computations that satisfy the privacy requirements of the law, as well as a set of computations that it considers to fail to provide sufficient privacy protection. In addition, if such an interpretation is not specified with precise, mathematical language, then there are most likely also computations that fall into a gray area and are neither unambiguously endorsed nor rejected by that interpretation. A conservative approach to modeling the law attempts to identify those computations that fall unambiguously within the intersection of all (or, at least, a large class of) reasonable interpretations of the law. That is, if a conservative model considers a certain computation to provide sufficient privacy protection, then all these reasonable interpretations of the law would also consider it to provide sufficient privacy protection.

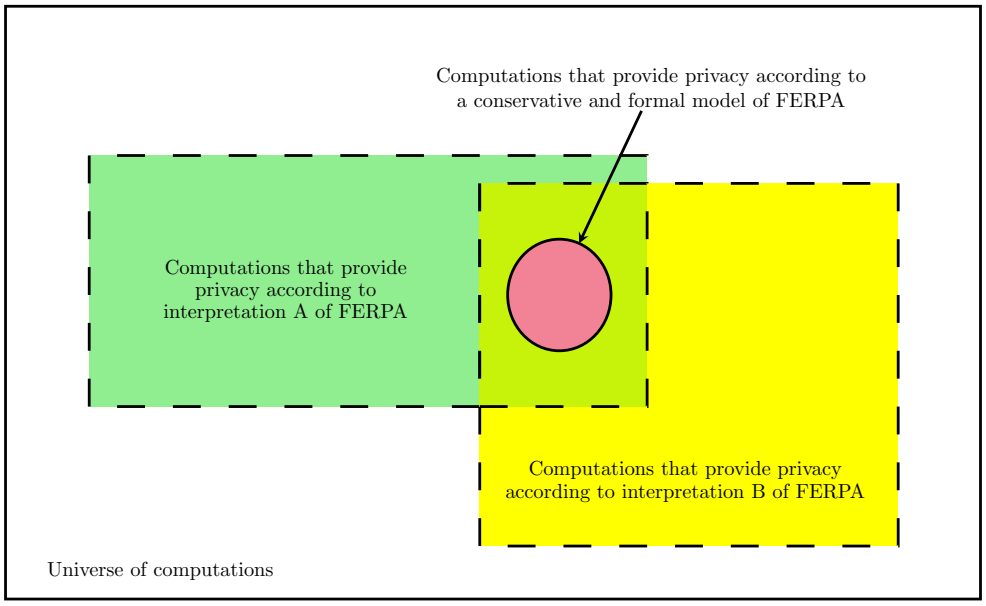

Figure 6: A conservative approach to modeling. The outer box delineates the universe of all computations. The inner boxes delineate computations that are considered to provide privacy protection according to particular interpretations of FERPA. We draw the borders of these boxes with dashed lines, since at the "edge" of a (non-formal) interpretation there are often computations that are not clearly accepted or rejected by that interpretation. A conservative approach to modeling attempts to identify those computations that fall safely within the intersection of all reasonable interpretations. In a formal model there is no ambiguity about whether a computation provides privacy by its standards, so we can draw its border with a solid curve.

In the following sections we describe the model that we have developed to capture the privacy

desiderata of FERPA's regulators. While at many points in the modeling process we made conservative decisions, we should also note that there are places in our model where we make some assumptions that are not fully conservative. We plan to extend our analysis to reflect a more fully conservative interpretation in future work. As we describe our model we note the places where we make assumptions that are not fully conservative.

4.2 Modeling FERPA's implicit adversary

FERPA does not specify an explicit model of the adversary its protections are intended to withstand. For instance, neither the statutory nor regulatory text specifies the capabilities of the hypothetical attacker who is trying to learn personally identifiable information held in education records and whom measures taken to ensure the privacy of this information must prevent from learning this information.[206] Despite the lack of an explicit description of the adversary envisioned, the Department of Education provided some details relevant to determining the types of attacks and attackers that were considered when drafting the regulations. In particular, FERPA's definition of *personally identifiable information* describes what or whom the law is designed to protect against. In this section, we argue that this definition and how it has been interpreted in agency guidance provides useful details that can serve as a basis for modeling the implicit adversary the Department had in mind when formulating FERPA's requirements.

As discussed in detail above in Section 3.2, FERPA prohibits the disclosure of non-directory personally identifiable information from education records, except with the consent of the student or her parent or guardian, or in accordance with one of the limited exceptions set forth by FERPA.[207] We might naturally ask how the agency originally envisioned an improper disclosure. In our investigation of this question, we are particularly interested in the case in which a school or educational agency releases information pursuant to the provision of FERPA permitting the release of de-identified data.[208] This exception provides that

> [a]n educational agency or institution, or a party that has received education records or information from education records . . ., may release the records or information without the consent required . . . 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.[209]

To qualify under this exception, the released data must not contain non-directory personally identifiable information. As discussed in Subsection 3.2.3, FERPA defines personally identifiable information to include direct and indirect identifiers, as well as

[206] For instance, the Department of Education declined to interpret the reference to a "reasonable person in the school community" within FERPA's definition of personally identifiable information as restricting the technological or skill level of a potential adversary. *See* 73 Fed. Reg. at 74,831 ("The standard proposed in paragraph (f) [of the definition of personally identifiable information] regarding the knowledge of a reasonable person in the school or its community was not intended to describe the technological or scientific skill level of a person who would be capable of re-identifying statistical information or redacted records.").

[207] See 34 C.F.R. § 99.30.

[208] 34 C.F.R. § 99.31(b)(1).

[209] *Id.*

> information that, alone or in combination, is linked or linkable to a specific student that would allow *a reasonable person in the school community, who does not have personal knowledge of the relevant circumstances*, to identify the student with reasonable certainty.[210]

By the inclusion of the quoted language, the agency emphasized concerns that a member of the school community might be able to learn non-directory personally identifiable information about a student from a disclosure of de-identified data, and established this as a category of privacy breach to protect against. As described above in Subsection 3.2.1, this type of privacy breach is relevant to the use of formal privacy models. Therefore, we take the "reasonable person in the school community, who does not have personal knowledge of the relevant circumstances" to be the implicit adversary embedded within FERPA's requirements for the purposes of our model.

Next, we explore who is considered to be a "reasonable person in the school community," and the knowledge such an individual is expected to have. The preamble to the 2008 final rule updating the FERPA regulations provides some limited guidance on these questions. As noted in Section 3.2, a "reasonable person" is described in the preamble to the amendment as a "hypothetical, rational, prudent, average individual."[211] In addition to enjoying the insider information about students that comes from being a member of the "school community," this individual is "also presumed to have at least the knowledge of a reasonable person in the local community, the region or State, the United States, and the world in general."[212] Moreover, the agency expressed an intent for this standard to provide "the maximum privacy protection for students."[213] At the same time, since the adversary is assumed not to have "personal knowledge of the relevant circumstances," we conclude that the adversary also has some uncertainty about the student information. In other words, the regulations appear to recognize that it is not necessarily a privacy breach if some private information is identified in a data release. Rather, it amounts to a privacy breach only if the adversary had some uncertainty about that information before the data were made public.

4.3 Modeling the adversary's knowledge

The regulatory language suggests that a "reasonable person in the school community" brings with her some knowledge about student information. We seek to model the adversary's knowledge in the privacy game we construct. How we model the adversary's knowledge will affect the adversary's probabilities of correctly guessing the private information of a student both with and without access to the result a computation performed on that private information.[214]

[210] 34 C.F.R. § 99.3 (emphasis added). FERPA's definition of personally identifiable information also includes "[i]nformation requested by a person who the educational agency or institution reasonably believes knows the identity of the student to whom the education record relates." *Id.* We do not extensively address this part of the definition in our privacy model. We note, however, that the model we derive for FERPA guarantees that, no matter the adversary's *a priori* knowledge about a student, the adversary's ability to guess private information about that student does not improve much by seeing the result of a computation that meets our definition of privacy. Thus, even if the adversary believes that an aggregate statistic reflects private information of a particular student, the statistic will not enable the adversary to guess private information about that student with much greater success than if the adversary did not have access to the statistic.

[211] 73 Fed. Reg. 74,806, 74,832 (Dec. 9, 2008).

[212] *Id.*

[213] *Id.*

[214] Note that, while having more knowledge makes it easier for the adversary to win the privacy game, having more knowledge also increases the baseline probability of success without access to the computation. While the former

We choose to model the adversary's knowledge about students as probability distributions over student attributes.[215] We believe that probability distributions provide a sufficiently rich language to model the type of knowledge that the regulations anticipate an adversary could have about students in protected education records. For instance, distributions can describe statistics that the adversary knows about the entire student body (e.g., demographic information), as well as beliefs that the adversary might hold about a particular student. For each student, we presume that the adversary has some *a priori* beliefs about that student, which we represent as a probability distribution. That is, in our model, each student is associated with a probability distribution that represents the adversary's beliefs about the non-directory personally identifiable information of that student.[216] The following examples demonstrate the versatility of probability distributions for modeling the adversary's knowledge.

First, probability distributions can be used to describe an adversary's general knowledge about the demographics of a school district. Suppose there is a school district with two schools, Abington and Bragdon. The percentage of low-income students at Abington is 95%, while only 40% of the students at Bragdon fall into this category. Furthermore, say that 35% of the students in Abington and 15% of the students in Bragdon scored below proficient on a statewide math exam. Without knowing anything else about her, if the adversary knows that Alice attends Abington, he might believe that there is a 95% chance that Alice is from a low-income family and a 35% chance that she scored below proficient on the exam. On the other hand, the fact that Grace attends Bragdon might lead the adversary to believe that there is a 40% chance that she is from a low-income family and a 15% chance that she scored below proficient on the exam. We can model this general knowledge with two distributions. A student sampled randomly from the first distribution has a 95% chance of coming from a low-income family and a 35% chance of scoring below proficient on the assessment. A student sampled randomly from the second distribution will fall into each of these categories with a likelihood of 40% and 15%, respectively.

Furthermore, distributions can be tailored to reflect more complex beliefs about trends across demographics. For instance, the two distributions described above could be reworked to reflect the adversary's belief that a student from a low-income family has a much greater chance of scoring below proficient on the examination than a student who is not from a low-income family. For example, consider the distributions given in Table 1. These distributions still reflect the fact that 95% of the students at Abington are low-income and 40% of the students at Bragdon are low-income, and that 35% of students at Abington scored below proficient on the exam and 15% at Bragdon scored at this level. However, the distributions now also reflect the fact that, if a student (from either school) is from a low-income family, that student scored below proficient on the exam with a likelihood of about 35%. Similarly, if a student is not from a low-income family, the likelihood

makes violating privacy easier (presumably, having more knowledge makes it easier to guess the private information of a student), the latter makes violating privacy harder (the baseline success probability is higher and, therefore, so is the threshold for a privacy breach). It follows that an adversary having more knowledge does not necessarily make protecting privacy harder. To see why this (somewhat counterintuitive) argument holds, consider the extreme case in which an adversary has complete knowledge of all the students' non-directory personally identifiable information. The adversary's successful guess of the private information of a student after seeing the outcome of a computation should not in this case be considered a failure of the computation to preserve privacy.

[215] Intuitively, a probability distribution describes the properties of random members of a population. Probability distributions are often used to model uncertainty in elements of a population.

[216] We assume that student attributes are drawn independently from the distributions; that is, the fact that one student has certain attributes does not affect the probability of another student having particular attributes. This is a limitation of our model, and will be discussed later in this section and in Section 4.12.

	Distribution describing student from Abington		Distribution describing student from Bragdon	
	Low-income	Not low-income	Low-income	Not low-income
Proficient or higher	60.1%	4.9%	26.0%	59.0%
Below proficient	34.9%	0.1%	14.0%	1.0%

Table 1: Distributions describing students at Abington and Bragdon. Numbers are given as probabilities. For instance, if we were to sample randomly from the distribution describing a student from Abington, the sampled student scored at least proficient on the exam and is from a low-income family with a probability of 0.601. On the other hand, the probability that the sampled student scored below proficient and is not from a low-income family is only 0.001.

that the student scored below proficient is only about 2%.

Second, in addition to describing beliefs based on demographic information, distributions can also reflect an adversary's more specific beliefs about individuals. This can be used to model the case in which the adversary has a substantial amount of outside knowledge about a particular student. For instance, the adversary might have heard that a student, Ashley, has a fiery temper. This knowledge might lead the adversary to believe that there is a greater probability that Ashley has been disciplined at school than the average student. Additionally, say that the adversary is aware that Ashley's mother is a professor of mathematics. Because of her family environment, the adversary might believe that Ashley is more proficient at math than the average student and so has a greater chance of having passed the state math exam. We can model the adversary's beliefs by associating with Ashley a distribution that describes a student that has a higher than average probability of having been disciplined and a higher than average probability of having passed the math exam.

One important limitation of our current modeling is that we assume that the characteristics of a given student are independent of the attributes of all other students. This means that we do not model *correlations* between students. We might expect to see such correlations in the real world. For example, there might be correlations between a small number of students, such as siblings. If one sibling comes from a low-income family, then all the other siblings should also have this characteristic. The best we can currently do in our model is to give each sibling an equal probability of coming from a low-income family. There might also be correlations among larger groups of students. For example, it might be the case that it is known that Alfred was a top achiever in his class on the state standardized math exam. Alfred's personal attribute (i.e., his exam grade, or whether he passed the exam) is not independent of the grades of the other students in his class. His grade is at least as high as the exam grades of the other students. More precisely, it is at least as high as the average, median, and 99th percentile grade in his class. While our current model does not account for correlations between students, in Section 4.12 we discuss some ways that we could address this limitation in future work.

4.4 Modeling the adversary's capabilities and incentives

As discussed in Section 3.2.3, the preamble to the 2008 final rule to update the FERPA regulations interprets the regulatory language with respect to the capabilities of a potential adversary. Specifically, the Department of Education explained that the "reasonable person" standard "was

not intended to describe the technological or scientific skill level of a person who would be capable of re-identifying statistical information or redacted records."[217] From this, we understand the intent was to make no assumption about an adversary's capability to re-identify students from "de-identified" data. Accordingly, we do not assume anything about the skill level of the adversary in our model. We also make no assumptions about the analyses the adversary can perform nor about the computational resources available to her.[218] Furthermore, we make no assumptions about the motivation of the adversary. By not constraining the resources available to the adversary or the adversary's motivation, our model conservatively accounts for the types of adversaries contemplated by many reasonable interpretations of FERPA.[219]

We also greatly expand our model by treating the adversary as representative of a whole class of potential attackers and attacks, rather than restricting our framework to only a particular attacker or attack. In fact, our model accounts even for attacks that have yet to be conceived. This approach is consistent with our commitment to adhering to a very conservative interpretation of FERPA, and it is also well supported by modern cryptographic theory. Modern cryptography emphasizes the design of cryptographic systems that are secure not only against known attacks, but also against attacks that were unknown at the time the cryptographic system was designed. Before modern cryptography's emphasis on providing robust protection against known and unknown attacks, cryptographic techniques typically only addressed known privacy threats. The failure of these cryptographic methods to provide privacy against new attacks motivated the development of new cryptographic techniques, but these techniques in turn were broken by even more sophisticated attacks. To break this cycle, modern cryptographers strive to develop methods that are guaranteed to be invulnerable to any attack that is feasible under a general computational model. By not making assumptions about the capabilities of the adversary in our modeling (i.e., beyond adhering to a general computational model), we ensure that our model is robust not only to currently known privacy attacks, but also to attacks developed in the future, as is consistent with a modern cryptographic approach.[220]

This approach is also consistent with the preamble to the 2008 final rule, which recognizes that the capabilities of attackers are increasing over time. For instance, the Department of Education recommends limiting the extent of information publicly released as directory information, with the justification that "since the enactment of FERPA in 1974, the risk of re-identification from such information has grown as a result of new technologies and methods."[221] Furthermore, in explaining why they left ambiguity in the definition of personally identifiable information, the regulators emphasize that holders of education data must consider potential adversaries who may attempt to learn private student information through many possible attack vectors.[222] The regulators contrast

[217] 73 Fed. Reg. at 74,831-32.

[218] We note, however, that the our modeling does not focus on the security of the particular implementation. In other words, we assume that the adversary only tries to learn private information from a data release. We are not concerned in this work with an adversary who might steal a hard drive containing sensitive data, or hack into the system to inspect its memory or actively manipulate it. We believe that this modeling is consistent with the privacy threats addressed by the regulations in the context of releasing de-identified information.

[219] For more detail on this approach, recall our conservative modeling desideratum illustrated in Figure (6).

[220] In particular, the adversary in our model may possess greater computational power and perform more sophisticated computations than those performed by the system designed to protect privacy or those conceived by the system designer. This is justified, in part, considering that once a system is put to use and the results of its computations are made public, an attacker may have virtually unlimited time to exploit it, applying in the process technological advances not known at the time of the system's design and deployment.

[221] 73 Fed. Reg. at 74,834.

[222] "It is not possible, however, to list all the possible indirect identifiers and ways in which information might

this approach with the HIPAA Privacy Rule, which effectively assumes that an adversary will only attempt re-identification attacks and base these attacks on a small set of a person's attributes.[223] Additionally, while traditional re-identification attacks against microdata are emphasized, FERPA does not contemplate only this type of attack, and implementation guidance also addresses exploits aimed at learning private student information from aggregate data.[224] Because the Department of Education has not explained the full scope of the types of attacks covered by FERPA, a conservative approach ensures our model accounts for *any* attack that can be perpetuated by an adversarial observer.

In contrast with our approach, some frameworks for measuring the privacy risk associated with a data release explicitly require making assumptions about the motives and capabilities of a potential adversary. For instance, some experts suggest that estimates of the motives and capabilities of a potential adversary should be used as an input when calculating the re-identification risk of a data release. If it is believed that any potential adversary will have limited resources to apply towards a privacy attack or little incentive to attempt an attack, the risk of re-identification is assumed to be small, so the data can presumably be safely released with fewer protections, or with administrative controls in place.[225] Such an approach has clear benefits to utility as data can be released with little or no modifications, and it may be justifiable in specific settings, such as where the payoff of a successful attack on privacy can be demonstrated to be significantly lower than the cost of the attack. However, the reliance on assumptions regarding the privacy adversary may result in future susceptibility to attacks as incentives and analytic capabilities change over time. We take the more conservative stance that the adversary is fully capable of any privacy attack and fully incentivized to attack. This is necessary in order to capture a broad class of possible interpretations of FERPA, in light of the Department of Education's explanation that FERPA's privacy requirements are intended to provide "the maximum privacy protection for students,"[226] and its recognition that "the risk of re-identification may be greater for student data than other information because of the regular publication of student directories, commercial databases, and de-identified but detailed educational reports by States and researchers that can be manipulated with increasing ease by computer technology."[227]

4.5 Modeling student information

Like FERPA, we distinguish between directory information and non-directory personally identifiable information in our model. Because directory information can be disclosed in accordance with FERPA, our model assumes that the adversary has access to it.[228] Non-directory personally iden-

indirectly identify a student." 73 Fed. Reg. at 74,833.

[223] "Further, unlike the HIPAA Privacy Rule, these regulations do not attempt to provide a 'safe harbor' by listing all the information that may be removed in order to satisfy the de-identitification requirements in § 99.31(b)." 73 Fed. Reg. at 74,833.

[224] *See, e.g.,* National Center for Education Statistics, Statistical Methods for Protecting Personally Identifiable Information in Aggregate Reporting, SLDS Technical Brief: Guidance for Statewide Longitudinal Data Systems (SLDS) (Dec. 2010), https://nces.ed.gov/pubs2011/2011603.pdf.

[225] *See* Khaled El Emam & Luk Arbuckle, Anonymizing Health Data ch. 2 (2013).

[226] 73 Fed. Reg. at 74,832.

[227] *Id.* at 74,834.

[228] Note that, in reality, some schools make their directory information available only to certain members of the school community, such as teachers and parents, not the general public. However, consistent with a conservative approach, we assume it is available to the adversary because in many cases a potential adversary will have access to this information. For instance, the adversary's target may be a student at a particular school that shares its directory

tifiable information is private information generally not available to the adversary, although the adversary might hold some *a priori* beliefs about it. As mentioned in the previous subsection, we model these beliefs via probability distributions over student attributes. Therefore, in our model each student is associated with a record consisting of two pieces of information: a set of concrete values that constitutes the student's directory information, and a probability distribution describing the adversary's beliefs about that student's private attributes. In Section 4.7.1, we discuss how we model ambiguity concerning which student attributes are considered directory information and which attributes are considered non-directory personally identifiable information.

4.6 Modeling a successful attack

FERPA prohibits the non-consensual disclosure of non-directory personally identifiable information. According to the regulations, disclosure "means to permit access to or the release, transfer, or other communication of personally identifiable information contained in education records by any means, including oral, written, or electronic means, to any party except the party identified as the party that provided or created the record."[229] Therefore, outside of a disclosure according to an exception to FERPA, a data release by an educational agency or institution should not communicate any non-directory personally identifiable information to a recipient or user of the data it discloses. In our model, we formalize this by saying that the adversary wins the privacy game if she is able to correctly guess non-directory personally identifiable information of a particular student after seeing the output of a computation performed on the private data.

More precisely, the adversary wins the game if she successfully guesses a function of the private information of a student. This is a more conservative approach than only considering the adversary to have won the game if the adversary guesses the exact private information. This model recognizes that often learning (or inferring) something about private information could be a breach of privacy, even if the private information itself is not learned directly or completely. For instance, consider a scenario in which the private information of a particular student is his test score, which happens to be a 54 (graded on a scale from 0 to 100). If the adversary is able to learn from a data release that the student failed the test, we consider the adversary to have won the privacy game, even if the adversary has not learned the student's exact numerical score.

The FERPA regulations require a data release to preserve uncertainty about private student information.[230] In addition, as discussed in Section 3.2.3, many states prohibit reporting that 100% of students achieved certain performance levels. This recommendation is presumably based on the assumption that there is likely a reasonable person in the school community who does not know the academic performance of every student in the school, as is often the case. Reporting that 100% of students achieved a certain performance level removes any uncertainty from that individual's mind about each student's individual performance; therefore, non-directory personally identifiable information has been communicated.

For a more nuanced example, consider the release of a table that contains information about

information with the public, or the adversary may be a parent at a school that makes its directory information available to parents and teachers.

[229] 34 C.F.R. § 99.3.

[230] *See* 34 C.F.R. § 99.3 (defining personally identifiable information to include "[o]ther information that, alone or in combination, is linked or linkable to a specific student that would allow a reasonable person in the school community, who does not have personal knowledge of the relevant circumstances, to identify the student *with reasonable certainty*").

	Test score levels			
	Below basic	Basic	Proficient	Advanced
Native English speaker	4	6	11	5
English language learner	5	4	1	0

Table 2: A release of non-directory personally identifiable information prohibited by FERPA.

student test scores, such as Table 2. The table distinguishes between native English speakers and English language learners, and gives the number of students from each category who received scores at various levels, including below basic, basic, proficient, and advanced. Suppose the table reports that one English language learner achieved a proficient score on the exam and the other nine English language learners scored at either a below basic or basic level. This would constitute a disclosure of non-directory personally identifiable information. To see why, consider that the sole English language learner, Monifa, who achieved a proficient score, could learn from the publication of this table that her peer English language learners, such as her friend Farid, scored at one of the lower levels. Prior to the release of the table, Monifa presumably did not know the scores of her classmates. Thus, if we take the "relevant circumstances" to be the performance of her peers, she did not have "personal knowledge of the relevant circumstances." However, after the release, she can infer with certainty that each one of the other language learners, like Farid, performed below a proficient level on the test.[231]

Guidance on de-identifying data protected by FERPA also suggests that very low or high percentages should be "masked" when reported.[232] For example, consider a statistic reporting what percentage of a class of 41 students scored at the proficient level on an exam, where only one student, Siobhan, scored at the proficient level. Guidance from the National Center for Education Statistics recommends reporting that $\leq 5\%$ scored at a proficient level, instead of the true percentage, which is around 2.5%.[233] If the true percentage were reported, Siobhan would learn that all her peers failed to achieve this level, since she knows that she accounts for the entire 2.5% of proficient students. However, if the "masked" percentage is reported, Siobhan would no longer be able to reach this conclusion with certainty, since another student could have scored at her level (i.e., 5% of 41 is slightly more than 2 students out of 41.). As there is only a 1/40 probability that a given student besides Siobhan scored at a proficient level, Siobhan might be able to form a strong belief about the performance of each of her peers; however, she has not learned with certainty the performance level of any particular classmate based on this publication. Consequently, the data release is considered to preserve uncertainty about private student information.[234]

To fully capture this in our model, we need to preserve uncertainty about non-directory personally identifiable information in a data release. We do so by requiring a stronger property. We say that a system provides privacy protection only if gaining access to the system does not change

[231] This example is adapted from Example 1 in National Center for Education Statistics, Statistical Methods for Protecting Personally Identifiable Information in Aggregate Reporting, SLDS Technical Brief: Guidance for Statewide Longitudinal Data Systems (SLDS) 5 (Dec. 2010), https://nces.ed.gov/pubs2011/2011603.pdf.

[232] *See* National Center for Education Statistics, Statistical Methods for Protecting Personally Identifiable Information in Aggregate Reporting, SLDS Technical Brief: Guidance for Statewide Longitudinal Data Systems (SLDS) (Dec. 2010), https://nces.ed.gov/pubs2011/2011603.pdf.

[233] *See id.*

[234] *Id.* at 27-29.

any adversary's ability to successfully guess the private information of any student very much. In our model, an adversary holds some *a priori* beliefs about the non-directory personally identifiable information of each student. Based on these beliefs alone, the adversary can make a guess about the private information of a student. We say that a system provides privacy protection if the adversary's chances of guessing the private information correctly after seeing a data release are about the same as his chances of guessing correctly based only on his *a priori* beliefs. Specifically, we allow the adversary's success probability to change by a multiplicative factor that is constrained within a small range.[235] For instance, we might say that the adversary's probability of success should only change by a multiplicative factor of between 0.9 and 1.1. In this case, if some adversary is able to correctly guess that a particular student failed an exam with a probability of 0.6 based on his *a priori* beliefs about that student, he would be able to successfully guess that the same student failed the exam with a probability of no more than $0.6 \times 1.1 = 0.66$ after seeing the output of a privacy-providing computation. His chance of successfully guessing the performance of that student has changed, but the change is bounded. Our approach also preserves uncertainty when the initial uncertainty is small. If the adversary can initially guess that a student failed with a probability of success of 0.99, we guarantee that the output of a privacy-providing computation would not increase his chances of guessing correctly beyond a probability of success of 0.991. Intuitively, this is because the probability of success of the "complementary adversary" (who is interested in the probability that the student did not fail the exam) cannot change significantly.[236]

4.7 Towards a FERPA privacy game

In the next subsection we describe a *privacy game* and a corresponding *privacy definition* that fits the model we have extracted based on FERPA's requirements for protecting privacy in releases of education records. Before we do so, we need to tie up two loose ends in our model. First, earlier in this section, we described how student information is classified as directory information or non-directory personally identifiable information. However, we did not describe the contents of the student information. Second, in the previous subsection we made the argument that the adversary wins the game if he is able to correctly link a student's identity with a function of that student's sensitive information. However, there is always some baseline probability that the adversary is able to do so, even without receiving the output of the computation, and we must account for this fact as we attempt to define whether a computation provides privacy protection.

4.7.1 Accounting for ambiguity in student information

Before we can run the privacy game we have defined, we also need to define what constitutes directory information and non-directory personally identifiable information. In the game, the adversary will have direct access to the directory information and will also have some indirect knowledge

[235] More formally, this multiplicative factor is captured by the parameter ε (epsilon), which is typically a small constant. The adversary's belief is allowed to change by a factor of $e^{\pm\varepsilon}$. For a small epsilon, $e^{\varepsilon} \approx 1+\varepsilon$ and $e^{-\varepsilon} \approx 1-\varepsilon$. The epsilon parameter can be tuned to provide different levels of privacy protection.

[236] Given the same information, the "complementary adversary" makes the opposite guess of the original adversary. The "complementary adversary" has an initial chance of $1 - 0.99 = 0.01$ of successfully guessing that the student did not fail the exam. This probability of success can only change by a multiplicative factor between 0.9 and 1.1, which means that after observing the computation output, the "complementary adversary" will have a probability of success of at least $0.01 \cdot 0.9 = 0.009$. Returning to the original adversary, we get that its probability of success is at most $1 - 0.009 = 0.991$.

of the private information. The computation will use both types of information to produce some output.

There is a degree of ambiguity in the regulatory definitions of directory information and non-directory personally identifiable information. The regulations provide a non-exhaustive list of examples of directory information,[237] but each educational agency or institution is granted discretion in determining what to designate and release as directory information. Seeking generality, we do not make any assumptions in our model about the content of directory information. Instead, *we allow the adversary to decide what information is published as directory information.* While this may seem like an unintuitive and unrealistic modeling choice, as it might be unlikely that any adversarial agent would have this ability in the real world, this modeling decision helps establish a privacy requirement that will be robust both to a wide range of potential interpretations of FERPA and to different choices made within specific institutional settings in accordance with FERPA. This modeling also exhibits conceptual clearness that we believe is instrumental to understanding intricate concepts like privacy. To recap this modeling decision, we are effectively allowing the adversary to choose directory information that is the worst-case scenario for the privacy protection of the system. If we are able to prove that the system provides privacy protection even in this worst-case scenario, then we will have confidence that it provides privacy no matter what the directory information actually is.[238]

Similarly, there is uncertainty about exactly what information constitutes non-directory personally identifiable information. From the definition of personally identifiable information, we know that the system must protect information that "is linked or linkable to a specific student."[239] However, we do not want to make any assumptions about the capabilities of the adversary or the methods he might use to identify a student in released records that have been stripped of non-directory personally identifiable information. Indeed, the guidance on interpreting FERPA instructs against making any assumptions of this nature.[240] It is impossible to say what information falls into these categories. Furthermore, we do not want to make any assumptions about the auxiliary knowledge that the adversary may have about students. Accordingly, as with the directory information, we allow the adversary to have a say about the non-directory personally identifiable information of the student or students whom he is targeting. More precisely, for each student in the dataset, we allow the adversary to choose the distribution over student attributes that models that student's non-directory personally identifiable information.

For example, the adversary could choose the student information presented in Table 3. The directory information consists of the name, age, and ZIP code of three students. The adversary chooses concrete values for these attributes. The private attributes are whether a student has a learning disability and whether a student has been suspended in the last year. In our modeling, the adversary assigns a probability distribution that describes the likelihood of each combination of these attributes being true. In the table, we explicitly break down the distribution for each student. Note that the probabilities of the outcomes for each student sum to one, as is required for a probability distribution. Consider the third row in the table. As directory information, the adversary chooses that this student's name is Samuel Strudwick, that he is seventeen-years-old, and that he lives in ZIP code 00034. The adversary assigns to Samuel a distribution describing

[237] 34 C.F.R. § 99.3.

[238] Note that, by allowing the adversary to choose the directory information, we are enabling the adversary to more easily win the game, while at the same time raising the bar for what qualifies as a privacy breach. See *supra* note 214.

[239] 34 C.F.R. § 99.3.

[240] *See* 73 Fed. Reg. at 74,831.

Directory information			*Probability distribution over private information*			
Name	Age	ZIP code	Disability & suspended	Disability & not suspended	No disability & suspended	No disability & not suspended
Robi McCabe	18	00034	0.3	0.3	0.2	0.2
Launo Cooney	17	00035	0.2	0.2	0.3	0.3
Samuel Strudwick	17	00034	0.1	0.2	0.3	0.4

Table 3: Example student information chosen by the adversary. Name, age, and ZIP code are the attributes given as directory information. The private information for each student is represented by a probability distribution describing the probabilities of all combinations of the following two binary attributes: whether that student has a learning disability, and whether that student has been suspended in the last year.

how likely he is to have a disability and how likely he is to have been suspended. This distribution gives an exact probability that each combination of the two binary attributes is true. For instance, according to this distribution chosen by the adversary there is a 10% likelihood that Samuel has a learning disability and has been suspended in the last year, and a 20% chance of Samuel having a learning disability, but not having been suspended. This distribution reflects the adversary's *a priori* beliefs about Samuel's private information. Additionally, during the game, Samuel's actual private information will be sampled from this distribution.

4.7.2 Accounting for the adversary's baseline success

Recall the cryptographic privacy game from Section 2.2, in which one of two plaintext messages is encrypted. The adversary is given the resulting ciphertext, and must then identify which of the two original plaintext messages is behind the ciphertext. Since each message was encrypted with a probability of 0.5, the adversary can win the game with a probability of 0.5 without even examining the ciphertext, e.g., by guessing randomly between the two messages or by always choosing the lexicographically first message as the guess. This is the adversary's baseline for success. Even if a "perfect" cryptographic computation was used to encrypt the message, the adversary can still be expected to win the game 50% of the time.

Similarly, it is unreasonable to expect that the adversary will never win the FERPA privacy game that we propose, as there is always the possibility that the adversary guesses correctly by chance. For example, consider that Siobhan attends a school where 50% of the students scored below proficient on the state reading exam. Without any other knowledge, the adversary might guess that Siobhan scored below proficient on the test and have a 50% chance of being correct, provided that each student scored below proficient with an equal probability.

A system can still be considered to provide privacy protection even if the adversary wins the game with some likelihood (e.g., due purely to chance). To see why this is the case, consider a computation C that, on every input, outputs the result 0. Because the outcome of C does not depend on the input, it provides perfect privacy protection. Suppose that the adversary knows that 80% of the students in a school have a learning disability. If C is performed on the private student data, the adversary receives only the output 0, which provides her with no useful information for learning private information about the students, i.e., whether a particular student has a learning disability. However, if the adversary guesses that a given student has a learning disability, she will

win the game with a probability of 0.8. The fact that the adversary wins with a high probability should not be considered to be in contradiction to C providing perfect privacy protection, as the adversary could win the game with the same probability without access to the outcome of C.

To account for the adversary's baseline chance of success, we say that a computation provides privacy protection if the probability of the adversary winning the game when she has access to the system is not much greater than the probability of her successfully guessing sensitive information without having access to the system. This approach closely mirrors the approach taken in Section 3.1. By this standard, the 0-outputting computation C from the previous example is considered to provide perfect privacy protection, since the adversary's probability of winning the game has improved not one iota by having access to the output of this computation. Since there is no informational relationship between the input dataset and C's output, C necessarily cannot leak information about the input, and so seeing C's output cannot possibly enable the adversary to win the game with a higher probability than before seeing the output. Of course, C provides no utility; it is a completely useless computation. Useful computations will involve some meaningful relationship between the input dataset and the computation output, and so we cannot expect them to provide the same level of privacy protection as C (i.e., perfect privacy). Nonetheless, if the adversary's probability of winning the game after seeing the output of one of these computations does not improve very much relative to her probability of success before seeing the output, then intuitively the computation must not have leaked very much private student information, and so we consider that computation to preserve privacy.

4.8 The game and definition

This game represents a scenario in which the adversary is committed to attempting to learn non-directory personally identifiable information about a particular student. For instance, consider the case of a local journalist trying to learn the private information of a school's star basketball player from a data release about school disciplinary actions. The journalist only cares about learning the information of this one student.[241]

In this model of a targeted attack, the adversary commits to attacking a specific student *before* seeing the output of a computation performed on private student data, and then attempts to guess private information about only that student. We illustrate the privacy game for this scenario in Figure 7.[242]

4.8.1 Mechanics

In line with a conservative approach to modeling as described above in Section 4.1, we allow the adversary to choose a directory of public student information. The adversary assigns to each student a probability distribution that describes the adversary's *a priori* beliefs about the non-directory

[241] This scenario is related to subsection (g) of FERPA's definition of personally identifiable information, which includes "[i]nformation requested by a person who the educational agency or institution reasonably believes knows the identity of the student to whom the education record relates." 34 C.F.R. § 99.3. In both cases, the contemplated adversary is trying to learn information about one particular student. However, in our model the adversary does not request education records. Instead, the adversary observes the result of a statistical computation performed on education records, e.g., attempts to learn information about the star basketball player from a de-identified data release. See *supra* note 210.

[242] In Section B.1, we discuss how to model an untargeted attack scenario, in which the adversary chooses which student to attack after seeing the output of the computation.

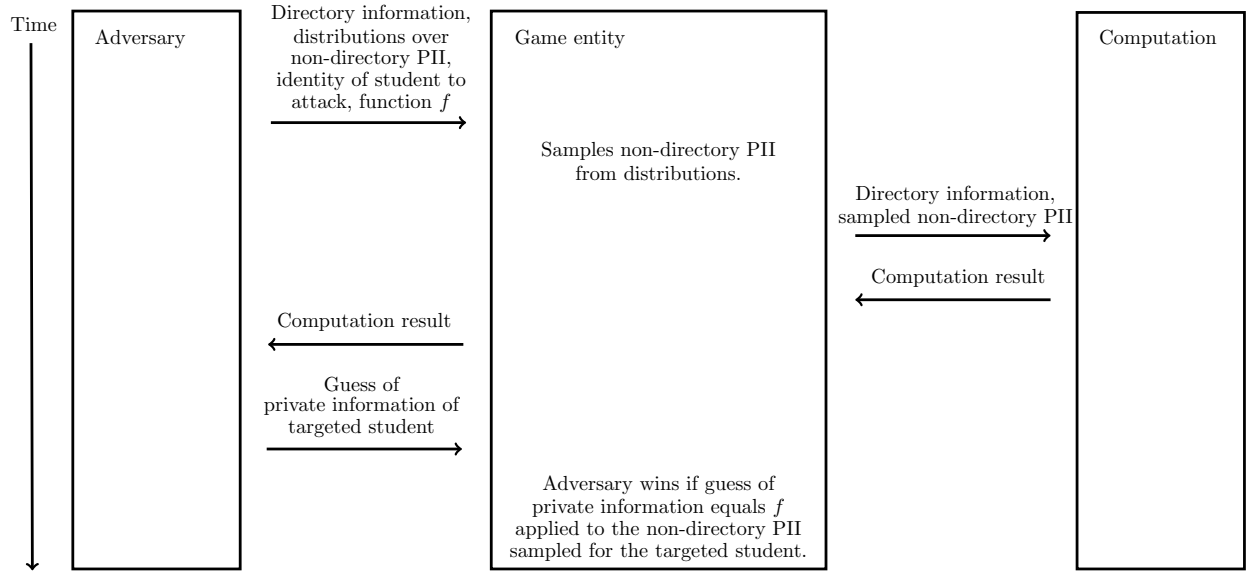

Figure 7: The real-world scenario.

personally identifiable information for that student, and also chooses a function f whose domain is private student information. Intuitively, f represents the aspect of private student information that the adversary is interested in learning. For example, consider a scenario in which the private information of a student being protected is his test score. The adversary might want to learn whether some student Bill passed or failed an exam, but might not care about the exact score that Bill earned. If this is the case, the adversary can choose a function *processScore* that takes Bill's numerical score and outputs "passed" if he earned a passing score and "failed" otherwise.

The adversary provides the directory, the probability distributions over student information, and the function f to the game mechanics. Additionally, the adversary chooses a particular student to attack (i.e., the adversary commits to trying to learn the non-directory personally identifiable information of that student, and that student alone), and provides the identity of the chosen student to the game mechanics. The game mechanics instantiate a database of non-directory student information by making a random draw from each of the probability distributions chosen by the adversary. This database, along with the directory information, is given to some computation C, resulting in an output statistic. Note that C is not supplied with the identity of the student that the adversary is attacking or the function that the adversary has chosen.

The game mechanics pass the result of C to the adversary. Based on this result, the adversary reports a guess about some aspect of the student's non-directory personally identifiable information. The game mechanics declare that the adversary has won if the guess matches the result of applying f to the distinguished student's non-directory personally identifiable information. Otherwise, the game mechanics declare that the adversary has lost. To continue the example from above, the adversary will guess either "passed" or "failed," and will win if his guess matches the result of applying *processScore* to Bill's test score.[243]

[243] The game mechanics' declaration of whether the adversary won or lost is only for purposes of the privacy definition (as a proxy for measuring the adversary's success rate) and does not correspond to an actual "real world" declaration. In fact, such a declaration would itself constitute a privacy breach (e.g., if the adversary has multiple

4.8.2 Privacy definition

Informally, we consider a computation to provide privacy protection in a targeted scenario if any adversary's chance of successfully winning the game against that computation is about the same as that of an adversary who does not have access to the output of the computation. We formalize this notion by comparing the game described above, which we call the real-world scenario, to another game, which we call the ideal-world scenario (see Figure 8). In the ideal world, the adversary chooses the same distributions over student information $\bar{P}$, directory information $\bar{d}$, student to target s, and function over student information f that the real-world adversary chose. In both scenarios the game mechanics instantiate a database of student information by sampling randomly from $\bar{P}$.[244]

However, unlike in the real world, in the ideal world no computation is performed on this database, and no information flows from the game mechanics to the adversary. Hence, the adversary clearly does not receive any private information about the targeted student during the course of the game, and can only guess based on his *a priori* beliefs about the student. Thus, as the name implies, the ideal-world scenario provides perfect privacy protection to the targeted student.[245]

It is important to note that, technically speaking, the ideal-world adversary and the real-world adversary represent distinct classes of adversaries. A real-world adversary makes a guess about the private student information based on seeing the computation result, whereas an ideal-world adversary guesses without receiving any information. We say that a computation provides privacy protection if for every real-world adversary playing against that computation, there is an ideal-world adversary who wins against that computation with nearly the same chance of success. As argued above, the ideal-world adversary learns absolutely nothing specific to the targeted student during the course of the game. If he is nearly as successful at guessing the private information of the student as the real-world adversary (who does see the computation result) then it must be that the computation result does not reveal significant information that it specific to the student to the real-world adversary. Thus, if for every real-world adversary there exists an ideal-world adversary with nearly the same level of success, we can be confident that the computation does not leak significant private information and hence provides privacy protection in the targeted setting.

chances to guess).

[244] The fact that the "actual" student attributes are drawn from distributions supplied by the adversary implies that the adversary necessarily has correct prior beliefs about the private information of the students. It would be too restrictive to guarantee that the computation output does not enable any adversary, regardless of its prior beliefs, to improve its chance of winning the game by very much, as this would effectively mean that the computation could not provide any utility. For instance, say that the private information is whether a student passed the exam and that the adversary (incorrectly) believed *a priori* that each of the students taking the exam had a 99% chance of failing, while in reality each had a 99% chance of passing the exam, and in fact 97% of them passed the exam. In addition, say the school wants to release the passing rate via some private computation. If a computation is not considered private if it improves this clueless adversary's chance of guessing whether or not a student passed the exam, then the school will not be able to release any useful statistic. However, we should note that, if the adversary could not guess with certainty the private information of a student before seeing the output of a computation that meets the definition of privacy in this section, it will not be able to guess with certainty the private information after seeing the computation output, no matter how good or bad its initial beliefs are. That is, uncertainty about student information is preserved by the computation.

[245] Note that the adversary can still win the game, despite the fact that no private information is leaked; see the discussion of baseline success in Section 4.7.2.

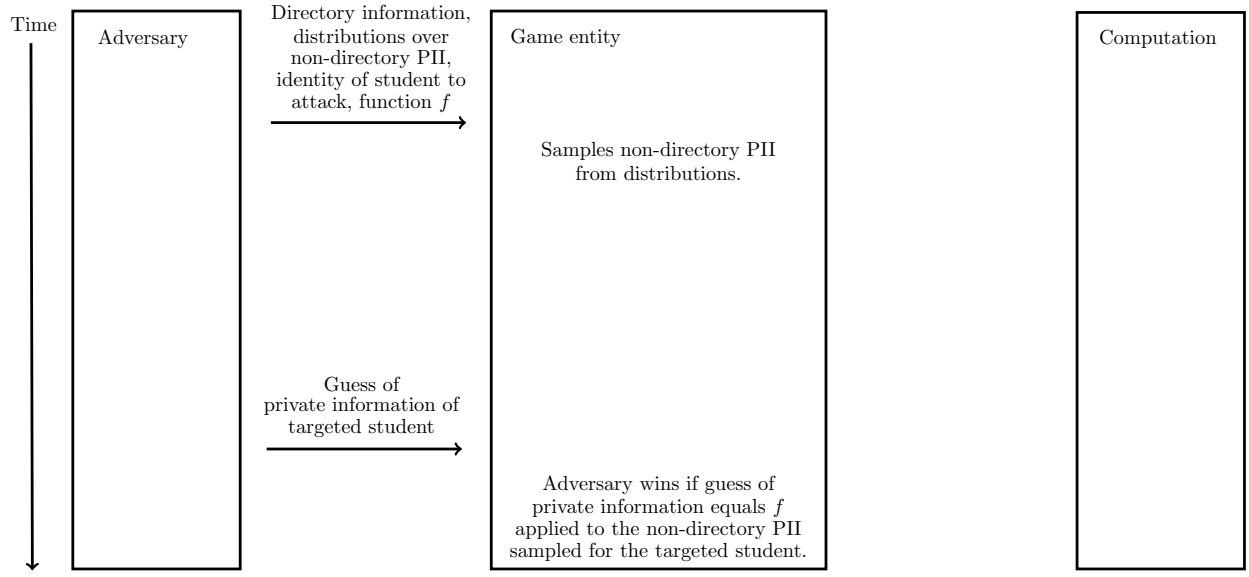

Figure 8: Ideal-world scenario. Note that the computation is left only for visual comparison with Figure 7. The computation could be removed from this figure as it is not applied to the data. Note further that there is no information flow from the game entity to the adversary.

4.8.3 The privacy loss parameter

In the above definition, we state that for every real-world adversary playing against a certain computation, there should be an ideal-world adversary with *nearly* the same level of success playing against that computation. We capture exactly how much better the real-world adversary is allowed to perform compared to the ideal-world adversary with a privacy-loss parameter. This parameter is a small multiplicative factor denoted ε. If the ideal-world adversary has success with probability p, the real-world adversary is only allowed to have success up to $p \times (1+\varepsilon)$. For instance, consider ε set to 0.01. If the ideal-world adversary's probability of success was 0.5, the real-world adversary's probability of success is only allowed to be $0.5 \times (1 + 0.01) = 0.505$. Thus, while the ideal-world adversary has exactly a 50-50 chance of winning the game in this example, the real-world adversary can be expected to win the game slightly more than 50% of the time.[246]

The fact that our definition is parameterized should be considered a strength, since it allows the

[246] It is important that this parameter is a multiplicative factor, and not additive. In the above example, an additive parameter would allow the adversary to win with a probability of $0.5+\varepsilon = 0.51$. While this might seem acceptable in this case, it is not acceptable when the ideal-world adversary's chance of winning is very small. For instance, say that the ideal-world adversary is guessing about some property that is extremely uncommon, such as one that only one out of ten thousand students has. The ideal-world adversary has a chance of success of $p = 1/10000 = 0.0001$. If epsilon were an additive parameter, then the real-world adversary's chance of success could be as high as $p + \varepsilon = 0.0101$. This means the real-world adversary could be over one hundred times as successful as the ideal-world adversary. The ideal-world adversary only has a one-in-ten-thousand chance of winning the game, but the real-world adversary wins the game more than one out of one hundred times. This would not be an acceptable level of privacy leakage if the attribute in question were highly sensitive. Fortunately, since ε is multiplicative and not additive, any computation meeting our privacy definition guarantees that a real-world adversary in this example could not win the game with a probability of more than $p \times (1 + \varepsilon) = 0.000101$, which is only slightly more than the ideal-world adversary's chance of success.

privacy guarantees of the definition to be tuned to an acceptable level. Such parameters are typically set after a societal-technical process during which the mathematical-theoretical understandings are matched with practical real-life experience, and then reviewed and adjusted periodically. Examples include the 5% confidence often used as an acceptable level of uncertainty in statistics and the bit-length of RSA encryption keys. Where included in a parameterized privacy definition, these parameters enable the definition to be tuned to meet ethical or legal standards of privacy protection.[247]

4.9 Applying the privacy definition

In this section we discuss a few examples of the privacy definition in action. The key to understanding these examples is that if a computation meets the privacy definition, then we have the guarantee that the real-world adversary is only marginally more successful at guessing private student information than the ideal-world adversary (who does not have access to the computation output at all). It also follows that the computation output will not enable any adversary to guess private student information with certainty (assuming, of course, that the adversary could not guess the information with certainty before seeing the computation result).

These strong guarantees automatically rule out certain classes of attacks. For instance, consider linkage attacks, in which an adversary is able to breach privacy by combining the computation output with an external database. Since the ideal-world adversary does not see the computation output, it cannot perform any sort of linkage attack. Because the real-world adversary can only be slightly more successful than the ideal-world adversary, whatever linkage attack the real-world adversary might attempt to perform when it sees the computation output cannot help the adversary guess private student information very much. More generally, this reasoning can be extended to make similar arguments for other categories of re-identification attacks as well.

We now refer back to a few examples from previous sections in which privacy was breached, and explain how privacy would be preserved in these situations if a computation that meets the definition of privacy from Section 4.8.2 is used. Recall the story of Monifa from Section 4.6, in which Table 2 (showing the number of students who scored at the below basic, basic, proficient, and advanced levels on a standardized test, divided by whether the students were native English speakers or English language learners) is released. Since the table shows that only one English language learner performed at a proficient level, and Monifa knows that she achieved this score, she can infer that all of her peer English language learners scored below proficient. Since she presumably did not know this before the release of the table, the release has violated the privacy of the other students.

We could model this situation by envisioning an adversary that has similiar knowledge to Monifa. This adversary has no uncertainty about the private attributes of Monifa, but has some uncertainty about the private attributes of the other students. Further, suppose this adversary accurately knows the *a priori* probability that each of Monifa's peer English language learners failed the exam. A computation that meets our definition of privacy comes with the guarantee that this adversary is not able to use the computation output to guess with certainty the private information of any of the students about whom it had initial uncertainty. For instance, the adversary would be able to

[247] In future work, we plan to explore paradigms based on legal standards and best practices to identify a range of values for an initial setting of the privacy-loss parameter that would be considered reasonable in terms of satisfying the requirements of FERPA and other regulations. For more on this topic, see the discussion of further technical research directions in Section 4.12 below.

guess the private information of Monifa with certainty, which makes sense, since the adversary is designed to model her. However, the adversary would not be able to guess whether Monifa's friend Farid passed the exam with much more success than the adversary's *a priori* beliefs would allow.

What this means is that even if Monifa had accurate *a priori* beliefs about the private information of her peers, if the table were released via a computation that meets our definition of privacy, she could not learn anything from the table that would enable her to guess the academic performance of one of her peers with certainty, unless she were able to do so before the release of the table.[248]

In Section 4.6 we also discuss a similar scenario, in which there is a class of 41 students in which only one student, Siobhan, scored at a profient level on an exam. If the statistic were released that 2.5% of the class achieved a proficient level, Siobhan would learn with certainty that her peers failed to achieve this level (since $0.025 \cdot 41 \approx 1$ and Siobhan knows that she scored at a proficent level). If, instead, the percentage were calculated using a computation that meets the definition of privacy from Section 4.8.2, Siobhan would not be able to come to this conclusion.

We can construct an adversary who models Siobhan's knowledge and apply our privacy definition. The adversary would have certainty about Siobhan's private information, as well as some beliefs about her peers' private information, but not know this information with certainty. The definition from Section 4.8.2 guarantees that the computation output would not enable this adversary to guess the academic performance of Siobhan's classmates with certainty. Thus, the real-world Siobhan would also not be able to guess her peer's academic performance with certainty, even after seeing the computation output.

4.10 Modeling summary

In the modeling process detailed above, we relied on language from the regulations and implementation guidance to construct a formal privacy game that captures a very conservative interpretation of the privacy desiderata of FERPA. Recall that, in order to address ambiguities in the regulatory language, we made choices that erred towards stronger privacy requirements. In addition, the model focuses on a targeted attack scenario, in which a hypothetical adversary seeks to recover non-directory personally identifiable information about a particular student. During the game, the adversary selects the directory information and distributions describing private attributes of the students that will be available as resources to be leveraged in the attack. The adversary wins the game if, after receiving the result of a computation performed on the student data, she is able to successfully guess a function of the non-directory personally identifiable information sampled for the targeted student.

In summary, this game allows us to state a precise, mathematical definition for determining whether a statistical computation meets the privacy requirements of FERPA as we have conservatively modeled them. In the next section, we discuss how it can be formally proven that any computation that is differentially private meets this definition.

[248] Note that if Monifa had very inaccurate prior beliefs about the likelihood of her peers passing the exam, the computation output might radically change her probability of successfully guessing whether or not a peer passed the exam. We allow this for the reasons described in *supra* note 244. However, Monifa would never be able to guess the performance of her peers with certainty, no matter her initial beliefs (unless she was certain before seeing the computation output).

4.11 Proving that differential privacy satisfies the requirements of FERPA

In the sections above, we developed a formal definition of privacy protection based on the privacy requirements of FERPA. We can proceed by proving mathematically that any computation that is differentially private meets this definition, and (since the requirements of this definition are likely stricter than that of FERPA) thus satisfies the privacy requirements of FERPA.

We provide a brief overview of the mathematical proof here, leaving a more technical sketch of the proof for Appendix A. Consider a computation that is being evaluated for its adherence to the definition of privacy extracted from FERPA. To show that it does, we can construct a proof using the real-world (Figure 7) and ideal-world (Figure 8) games presented in Section 4.8 above, which were based on conservative interpretations of FERPA's privacy requirements. Recall that in the ideal-world game, no computation is performed on the data and no information flows from the game entity to the adversary, while, in the real-world game, the adversary does get to see the computation result. The goal of the proof is to demonstrate that, for every real-world adversary that plays the game against the computation, there exists an ideal-world adversary that is nearly as successful at winning the game. If this is true, we conclude that the real-world adversary also cannot win the game significantly better than in the case where it does not get to see the computation result.

The proof follows a paradigm known as a *hybrid argument*,[249] which involves a progression of hybrid games (depicted in Figure 9 in Appendix A). The proof's progression begins with the ideal-world game (which is identical to that depicted in Figure 8 above) and ends with the real-world game (which is identical to that depicted in Figure 7 above), with intermediate "hybrid" games that share features with both the ideal- and real-world games. The argument demonstrates that there is little to no difference in privacy loss between every consecutive hybrid games in the progression, and hence *little privacy loss between the ideal- and real-world games*.[250] Intuitively, this argument can be likened to proving that the distance between two points A and B is small by demonstrating that it is possible to reach point B from point A by making only a small number of small steps.

Based on this proof, we can conclude that using differentially private computations to analyze educational records is consistent with a very conservative interpretation of FERPA. Therefore, organizations who wish to perform analyses on non-directory personally identifiable information using these tools, or to allow others to do so using their data, can do so with high confidence that the risk of thereby being found in violation of FERPA is very low.

4.12 Extending the model

The model discussed in this Article accounts only for adversaries who target a particular student in the data and see only the result of a single computation performed on the student data. In this section, we introduce two possible extensions to this model, including extensions for untargeted attack scenarios and scenarios involving multiple data releases. A more detailed, technical discussion of these extensions is provided in Appendix B.

In an *untargeted attack scenario*, the adversary does not commit to attacking a particular student, but rather decided on the student to attack after seeing the computation result. This

[249] This paradigm originated in Shafi Goldwasser and Silvio Micali, *Probabilistic Encryption*, 28 Journal of Computer and Systems Science 270 (1984).

[250] The definition is parameterized by epsilon, which controls how much better the real-world adversary is allowed to do than the ideal-world adversary. The use of this parameter enables the analysis tool to be tuned to reflect different levels of privacy protection.

attack scenario captures an adversary like a data broker, who wants to extract from a data release information about *any* student in the dataset, but does not initially have any particular student in mind. The *multiple-release scenario* reflects the fact that it is common for multiple statistics to be published about the same group of students, and there is the possibility that the releases in aggregate fail to preserve the students' privacy, even if each release in itself seems privacy-preserving.

These extensions of the model reflect more general privacy protection than the targeted, single-release scenario we focus on in this Article. By this, we mean that every computation that preserves privacy in the untargeted scenario or the multiple-release scenario also preserves privacy in the targeted, single-release scenario, but the inverse is not true.

We describe informally why this is so in the case of the targeted and untargeted scenarios. Although it *allows* adversaries to choose which student to attack after seeing the computation output, the untargeted scenario also accounts for adversaries who commit at the outset to attacking a particular student. Since this is exactly the class of adversaries that the targeted model contemplates, any computation that preserves privacy in the untargeted scenario also preserves privacy in the targeted scenario.

On the other hand, not every computation that preserves privacy in the targeted scenario preserves privacy in the untargeted scenario. For example, consider a computation that leaks information about a randomly chosen student. This computation might still meet the requirements of the targeted model, since the probability that the computation leaks information about the same student that the adversary chooses to attack might be sufficiently small. However, if the adversary is able to decide which student to attack after seeing the computation output, the adversary is able to take advantage of the leaked information and win the game with an unacceptably high probability by choosing to attack the student whom the computation compromised.

By introducing the extensions to the untargeted attack and multiple release scenarios we do not mean to imply that the single release model formalized in Section 4.8 is insufficient or incorrect. On the contrary, the legal analysis and technical arguments presented throughout Section 4 suggest that the single release model is likely much stronger than many interpretations of what is needed to comply with FERPA. We present these extensions for two reasons. First, these extensions do correspond to what may be privacy concerns in real-world uses of educational data, so it makes sense to consider them when examining privacy-preserving technologies to be used with such data. Second, although we do not do so in this article, it can be shown that differentially-private computations preserve privacy in either extension to the model, making the argument that the use of differential complies with the standard presented in FERPA more robust.

Further technical research is needed to extend the model to address situations in which student attributes are dependent, and to select an appropriate value for the privacy parameter based on a legal standard of privacy protection. As noted in Section 4.3, we treat the attributes of a student as being independent of the attributes of any other student. This does not fully capture reality, as sometimes the fact that one student has some attribute makes it more likely that another student has a certain attribute. For instance, if a student has a learning disability, it is more likely that his identical twin also has that disability. Our analysis could be modified to protect privacy when there is dependence between members of small groups of students. In addition to correlations between a few individuals such as family members, there could be global correlations among the students not accounted for by our model. In general, it is not possible to provide non-trivial inferential privacy guarantees when the adversary has arbitrary auxiliary information about correlations concerning

members of a dataset.[251] However, guarantees can be given when the adversary's knowledge about these correlations meets certain conditions.[252] Therefore, one possible direction for addressing this issue in our model would be to change the game mechanics to allow the adversary to construct limited types of correlations between students. In this case, we would still be able to give privacy guarantees when students are not independent, provided that the correlations between the students fit the ones allowed in the model. We intend to explore the suitability of this direction in future work.

In addition, both the privacy definition developed throughout Section 4 and differential privacy are parameterized, i.e., they both refer to a privacy-loss parameter ϵ that controls the level of privacy required (and, as a result, the accuracy of the computations that can be performed). Intuitively, ϵ corresponds to the requirement that risk to privacy should be kept "very small," and the question of setting ϵ (briefly discussed in Section 4.8.3) corresponds to quantifying "very small." The setting of ϵ is likely to be a dynamic process, in which the value is adjusted over time, depending on the understanding of the various factors affected by the real-world use of differential privacy. It is important, however, to begin this process, and the question therefore reduces to developing and implementing a methodology for the initial setting of the privacy-loss parameter. In future work, we intend to explore a methodology whereby an initial value for ϵ is selected to match the differential privacy guarantees to the intent underlying current best practices (regardless of whether policymakers' expectations were actually met by a given standard and how it has been interpreted in practice).

5 Discussion

On first impression, the gaps between the standards for privacy protection found in statutes and case law, on one hand, and formal mathematical models of privacy, on the other, may seem vast and insurmountable. Despite these challenges, this Article demonstrates a case where it is indeed possible to bridge between these diverging privacy concepts using arguments that are rigorous from both a legal and a mathematical standpoint. The main contribution of this Article is therefore its approach to formulating a legal-technical argument demonstrating that a privacy technology provides protection that is sufficient to satisfy regulatory requirements. In this section, we discuss the significance, as well as the policy implications, of this contribution.

5.1 Introducing a formal legal-technical approach to privacy

This Article's primary contribution is a rigorously supported claim that a particular technical definition of privacy (i.e., differential privacy) satisfies the requirements of a particular legal standard of privacy (i.e., FERPA). The argument consists of two components. The first part of the argument provides support for the claim that FERPA's privacy standard is relevant to performing differentially private analyses of educational records. This part builds on results from the technical literature showing that the release of aggregate statistics can leak information about individuals. It is also supported by a legal analysis of the text of the FERPA regulations as well as agency guidance on interpreting FERPA that pertains to releases of aggregate statistics.

[251] *See* Cynthia Dwork & Moni Naor, *On the Difficulties of Disclosure Prevention in Statistical Databases or The Case for Differential Privacy*, 2 Journal of Privacy and Computation 93 (2010).

[252] *See* Arpita Ghosh & Robert Kleinberg, *Inferential Privacy Guarantees for Differentially Private Mechanisms*, Working Paper (2016).

The second part of the argument involves an extraction of a formal mathematical model of FERPA. This analysis identifies and characterizes an adversary, formulates a privacy definition, and constructs a proof demonstrating that differential privacy satisfies the mathematical definition extracted from the regulation. Based on FERPA's definition of personally identifiable information, we construct a *detailed model of a privacy adversary*, or an attacker trying to learn private information from the system. This analysis is rooted in FERPA's conceptualization of a potential attacker as "reasonable person in the school community, who does not have personal knowledge of the relevant circumstances."[253] Using this language as a reference point, the model specifies the type of prior knowledge an adversary may have. The model also includes a specification of the information provided to the attacker from the privacy system, namely, the information designated by the school as directory information and the result of a particular statistical analysis. Finally, it defines the attacker's goal in terms of a violation of FERPA, i.e., identifying non-directory personally identifiable information about a student.

Next, following a well-established paradigm from the field of cryptography, we extract a game-based definition of privacy based on FERPA's requirements. This privacy game is a thought experiment that aims to capture what it means for a statistical computation to preserve privacy under (a conservative interpretation of) FERPA. More specifically, the game aims to model what it means for non-directory personally identifiable information to be protected from leakage. The game is formalized as an interaction between an adversary and a system that provides a statistical analysis, and this interaction is mediated by the mechanics of the game. To model the variety of settings in which a privacy-preserving mechanism may be challenged, we let the adversary choose the directory information that is available for each student, the prior beliefs the adversary possesses about the private information for each student, and the particular student to target in the attack. These features of the model are formalized within the game by having the adversary supply each of these pieces of information to the game mechanics. The mechanics of the game generate hypothetical educational records that reflect the directory information supplied by the adversary, as well as the adversary's beliefs about the non-directory personally identifiable information of the students in the school. Then, the game mechanics feed these records into the statistical analysis and supply the adversary with the result of the analysis. Finally, the game mechanics monitor whether the adversary is successful in guessing the non-directory personally identifiable information of the targeted student using the results of the analysis she receives.

Key features of our approach are its mathematical formalism and its conservative design choices. These aspects of the approach enable us to construct a proof that differential privacy satisfies the definition of privacy extracted from FERPA. This approach provides a very strong basis for arguing that differentially private tools can be used to protect student privacy in accordance with FERPA when releasing education statistics. Furthermore, it provides a rationale for asserting that the use of differential privacy would be sufficient to satisfy a wide range of potential interpretations of FERPA's privacy requirements, including very strict interpretations of the regulations. Furthermore, extensions to the privacy definition formulated in this Article are provided in Appendix B, which demonstrates the robustness of the argument that use of differential privacy is sufficient to satisfy FERPA with respect to an even wider range of interpretations of the regulations.

[253] 34 C.F.R. § 99.3.

5.2 Policy implications and practical benefits of this new approach

The analysis in this Article embraces a scientific understanding of privacy. In particular, it recognizes the importance of relying on formal, mathematical definitions of privacy. Adopting formal definitions is critical to ensuring that the privacy technologies being developed today will provide strong privacy protection, withstand future kinds of attacks, and remain robust over the long term despite the increasingly wide availability of big data and growing sophistication of privacy attacks. We argue that bringing mathematical formalism to the law holds promise for addressing ongoing challenges in information privacy law.

This Article advocates the application of a scientific understanding of privacy and mathematical formalism within the practice of information privacy law, as elements of a new regulatory regime that will provide strong, long-term protection of information privacy. When evaluating whether a technology provides sufficient privacy protection in accordance with a legal standard, we recommend that experts support their determinations with a rigorous analysis. In particular, we suggest that their determinations be based on a formal argument that is well-supported from both legal and technical perspectives and that their analysis should include a detailed description of a formal mathematical model extracted from the regulation.

In this way, a legal-technical approach to privacy can facilitate the real-world implementation of new privacy technologies that satisfy strong, formal definitions of privacy protection. A challenge to bringing formal definitions to practice in real-world settings is ensuring that the definitions that emerge in the mathematical study of privacy are consistent with the normative privacy expectations set in the law. Privacy regulations often rely on concepts and approaches that are fundamentally different from those underlying formal privacy models, and their requirements are inherently ambiguous. For instance, many information privacy laws have historically adopted a framing of privacy risks that is based on traditional and mostly heuristic approaches to privacy protection like de-identification. This potentially creates uncertainty and risk for practitioners who would seek to use tools that do not rely on traditional techniques such as de-identification that are seemingly better supported by the law. However, an approach to formally modeling legal requirements that is conservative and consistent with a broad range of legal interpretations can provide a high level of assurance that the use of a privacy technology is sufficient under a legal standard.

Arguments such as the one presented in this Article can help lower the barrier for adoption of technologies that move beyond these traditional conceptions of privacy, including privacy technologies which adhere to formal privacy models. In particular, given a formalization of a regulation's privacy desiderata as a mathematical definition, privacy researchers can make and substantiate claims that certain computations satisfy the requirements of the formalization and hence comply with the regulation. This approach can be used to overcome ambiguities in legal standards for privacy protection and be used to design implementations of new privacy technologies that can be shown to satisfy a legal standard of privacy protection with more certainty than is afforded by alternative approaches. In turn, applications of this approach can support the wider use of formal privacy models, which provide strong guarantees of privacy protection for individuals whose information is being collected, analyzed, and shared using technologies relying on such models.

5.2.1 Benefits for privacy practitioners

This approach has potential applications and benefits for government agencies, corporations, research institutions, regulators, data subjects, and the public. Organizations such as research uni-

versities and corporations currently manage large amounts of personal data that hold tremendous research potential. Many of these organizations are reluctant to share data that may contain sensitive information about individuals due to recent high-profile privacy breaches and the specter of legal liability. Many government agencies, most notably statistical agencies, are interested in adopting public-facing tools for differentially private analysis. However, before sharing sensitive data with the public, agencies typically must demonstrate that the data release meets relevant regulatory requirements and satisfies various institutional policies, such as those related to any applicable internal disclosure limitation review. The proposed methodology could be used in this case to demonstrate that an agency's use of a formal privacy model satisfies its obligations to protect the privacy of data subjects pursuant to applicable laws such as the Privacy Act of 1974[254] or the Confidential Information Protection and Statistical Efficiency Act.[255]

Similarly, corporations and research institutions that collect, analyze, and share statistics about individuals may seek to use differentially private tools or other tools based on formal privacy models, but wish to do so only with assurance that doing so will be in accordance with regulatory requirements for privacy protection. Formal modeling, especially when done conservatively, could enable actors within each of these sectors to begin using and sharing data with assurance that the risk of an administrative enforcement action is low. For example, a formal legal-technical approach could be applied towards satisfaction of the expert determination method of de-identifying protected health information in accordance with the HIPAA Privacy Rule. An expert applying a privacy-preserving technology could use this formal approach both to "determine[] that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information" and to "[d]ocument[] the methods and results of the analysis that justify such determination."[256] This approach can be used to provide documentation of compliance with other privacy laws as well, providing a high degree of confidence in justifying—both ex ante and ex post—that an organization's practices with respect to privacy-preserving analysis and publishing meet the requirements of the law.

In addition, regulatory agencies such as the Department of Education and corresponding state agencies often develop specific guidance on implementing privacy safeguards in accordance with regulatory requirements. Agencies could employ a legal-technical approach to privacy in the future to evaluate and approve the use of new technologies, based on a rigorous determination regarding whether they satisfy existing regulatory and statutory requirements for privacy protection. This approach could also benefits for data subjects, and the public at large, facilitating the use of tools that provide stronger privacy protection than data subjects have been afforded in many historical cases, as well as enable new and wider uses of data that carry benefits to society.

5.2.2 Benefits for privacy scholars

Precise mathematical modeling can also enhance our understanding of the privacy protection afforded by various laws and technologies. Consider a technology that offers privacy guarantees that are weaker than those provided by differential privacy (but potentially outperforming differential privacy in other respects, such as utility). An expert seeking to demonstrate that this technology

[254] 5 U.S.C. § 552a.
[255] 44 U.S.C. § 3501 note.
[256] *See* 45 C.F.R. § 164.514(b).

satisfies FERPA may choose to extract and justify a weaker model of FERPA than the one extracted in this Article. This can enable a comparison of the various models (and, consequently, multiple privacy definitions) that have been developed based on FERPA. Such a comparison makes it possible to understand, in precise terms, the privacy guarantees offered by each model, as well as the assumptions relied upon by each model. Because the privacy guarantees and assumptions of each model are made explicit, policymakers, scholars, and the public can better understand the tradeoffs of each model. This provides a basis for scholarly and public policy debates regarding the privacy guarantees that should be required from normative and technical perspectives and the types of assumptions that are reasonable in a privacy analysis. Feedback from the normative debates can, in turn, help inform practitioners constructing models for evaluating privacy technologies and provide more clarity to those developing the technologies themselves.

Comparing the models extracted from different regulations could inform our understanding of the ways in which regulations differ, including evaluating the relative strength of protection provided by different privacy laws. The large number of privacy regulations that are applicable based on jurisdiction and industry sector make regulatory compliance complex.[257] The high level of abstraction offered by the analytical approach in this Article has the potential advantage of simplifying the analysis of privacy regulation and its applicability to particular privacy technologies. This approach can serve as a new lens for examining and comparing privacy regulations. For instance, it could be used to identify essential features that regulations ought to share and be used to reform such regulations in the future.

Although this Article does not advocate any particular regulatory changes, we argue that evaluations of new technologies should be done rigorously so as to ensure their design and implementation adheres to the legal standards of privacy protection. This Article highlights a number of gaps between the current legal framework for privacy protection and recent advances in the scientific understanding of privacy. Updating information privacy laws based on a modern scientific approach to privacy could bring greater certainty and stronger privacy measures to practice. Using mathematical formalism, reforms can be made to specify requirements that are meaningful and accurate and yet permissible enough to promote the development and use of innovative privacy technologies. Furthermore, mathematical formalism can serve as a tool for guiding regulatory decisions based on rigorous conceptions of privacy. To begin to close the gap between privacy law and technology, we recommend that future regulations aim to define the objective of a privacy standard, rather than providing a list of permissible privacy protection technologies or certain identifiers that, if redacted, presumably render the data de-identified.[258] This shift in regulatory approach would enable practitioners and legal scholars to assess whether new technologies meet the goals provided by the regulations. This would, in turn, help ensure that new privacy technologies can be brought to practice with greater certainty that they satisfy legal standards for privacy protection.

[257] *See* 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 (2011).

[258] For further discussion and examples illustrating how regulatory requirements could be stated in terms of a general privacy goal, see Salil Vadhan et al., Comments to the Department of Health and Human Services Re: Advance Notice of Proposed Rulemaking: Human Subjects Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators, Docket No. HHS-OPHS-2011-0005 (Oct. 26, 2011), http://privacytools.seas.harvard.edu/files/commonruleanprm.pdf; Alexandra Wood et al., Comments to the Department of Health and Human Services, Re: Federal Policy for the Protection of Human Subjects Proposed Rules, Docket No. HHS-OPHS-2015-0008 (Jan. 6, 2016), https://www.regulations.gov/document?D=HHS-OPHS-2015-0008-2015.

5.3 Applying this approach to other laws and technologies

In this Article, we demonstrate that it is possible to construct a rigorous argument that a privacy technology satisfies a legal standard of privacy protection, using differential privacy and FERPA for illustration. However, in doing so, we do not argue that using differential privacy is *necessary* to satisfy FERPA's requirements for privacy protection, nor that privacy-preserving technologies other than differential privacy are insufficient to satisfy FERPA. We also do not claim that the model presented in Section 4 is the only possible model that can be extracted from FERPA. Rather, we believe there are several possible models that can be extracted from FERPA to support the use of various privacy technologies.

In future work, we anticipate extending the technical results presented in this Article to a general methodology that can be applied to privacy models other than differential privacy and regulations apart from FERPA. The computational focus of this analysis works at a level of abstraction in which many of the differences between particular regulations are likely to become less important. Achieving this level of abstraction, however, will require deepening our understanding of how different regulations lend themselves to the kinds of analyses we perform. It will also require extending our "toolkit," i.e., the collection of argument paradigms that can be used in making a claim that a differential privacy, or another formal privacy model, satisfies a regulatory standard of privacy protection.

Establishing a general methodology will require examining regulations that are different from FERPA in terms of the type of analysis that would be required. Two examples of regulations with privacy requirements that differ in significant ways from FERPA's include the HIPAA Privacy Rule and Title 13 of the U.S. Code. In this brief overview, we discuss some of the ways in which the privacy standards in these laws differ from that set forth by FERPA. These differences will likely require future modifications to our argument in order to achieve a generally applicable methodology.

The HIPAA Privacy Rule presents a unique set of challenges for modeling its requirements formally. The Privacy Rule sets forth two alternative methods of de-identification: a detailed safe harbor method and a method relying on expert determination. While FERPA includes a description of a privacy adversary, its knowledge, and its goal as a "reasonable person in the school community, who does not have personal knowledge of the relevant circumstances, to identify the student with reasonable certainty," neither the regulatory language nor relevant guidance for HIPAA includes such an explicit description of the envisioned adversary. Furthermore, HIPAA's expert determination method delegates the responsibility to confirm that the "risk of identification is very small" to a qualified statistician,[259] but neither the text of HIPAA nor guidance from the Department of Health & Human Services provides the form of reasoning an expert must perform to argue that the risk is small. For instance, where the guidance describes "principles for considering the identification risk of health information," it notes that such "principles should serve as a starting point for reasoning and are not meant to serve as a definitive list."[260] The provision of general principles meant to serve as a "starting point" rather than specific criteria to be met creates challenges for modeling the regulatory requirements with precision.

Other challenges arise when modeling privacy standards that lack detail, an example being the confidentiality provisions of Title 13 of the U.S. Code, the statute which establishes the authority

[259] 45 C.F.R. § 164.514(b).

[260] Office for Civil Rights, Department of Health and Human Services, *Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule* (2012).

of the U.S. Census Bureau and mandates the protection of data furnished by respondents to its censuses and surveys. The model we extracted from FERPA draws from a large collection of relevant language from the regulations and guidance from the Department of Education. In comparison, Title 13's privacy standard is succinct, providing only that the Census Bureau is prohibited from "mak[ing] any publication whereby the data furnished by any particular establishment or individual under this title can be identified."[261] Rather than delineating specific privacy requirements in regulations or guidance, the Census Bureau delegates much of the interpretation of the confidentiality requirements of Title 13 to the technical experts of its internal Disclosure Review Board. Modeling the privacy requirements of Title 13 would likely require heavy reliance on sources beyond the text of the statute, such as policies published by the Census Bureau and prior determinations made its Disclosure Review Board.

In future work, we intend to model additional laws such as the HIPAA Privacy Rule and Title 13, as well as other privacy technologies, and we advocate further application of this approach by other scholars and practitioners. The results of this research are anticipated to point to alternative modeling techniques and, ultimately, inform a more general methodology for formally modeling privacy regulations.

5.4 Setting the privacy parameter

As noted in Section 4, the privacy definition we extracted from FERPA is parameterized by a numeric value denoted ϵ. Any real-world implementation of a privacy technology relying on this definition from FERPA would require the appropriate value of ϵ to be determined. An understanding of the parameter ϵ can be informed, in part, by the definition of differential privacy, which is also parameterized by a value often denoted ϵ. As illustrated by the proof presented in Appendix A demonstrating that differential privacy satisfies the requirements of the definition extracted from FERPA, the parameter ϵ is used similarly in both definitions. As it relates to the differential privacy definition, the parameter ϵ intuitively measures the increase in risk to an individual whose information is used in an analysis. In order to provide strong privacy protection, the recommendation would be to keep ϵ "small." However, a smaller ϵ implies less accurate computations. Setting this parameter therefore involves reaching a compromise between data privacy and accuracy. Because this tradeoff has both technical and normative components, choosing a value of ϵ should also be based on both technical and normative considerations.

In future work, we plan to explore methodologies for setting ϵ. A setting of ϵ controls both the level of privacy risk and quality of analysis. We anticipate that our arguments with respect to setting ϵ in real-world settings will rely both on a quantitative understanding of this tradeoff and on legal-normative arguments for balancing these concerns. In particular, relevant legal and normative concerns may include policymakers' expectations in terms of this tradeoff between privacy protection and information quality, as well as the expectations of data subjects.

6 Conclusion

As new privacy technologies advance from the research stage to practical implementation and use, it is important to verify that these technologies adhere to normative legal and ethical standards of privacy protection. Being able to demonstrate that a privacy technology satisfies a normative

[261] 13 U.S.C. § 9(a)(2).

standard will be critical to enabling practitioners to adopt such tools with confidence that they have satisfied their obligations under the law and their responsibilities to the subjects of the data.

However, making a rigorous, substantive claim that the protection offered by a privacy technology meets a normative standard requires the development of a common language as well as formal argumentation. These features are essential to bridging the gaps between divergent normative and technical approaches to defining and reasoning about privacy. Towards this end, this Article details a combined legal-technical argument for bridging between the privacy definition found in a legal standard, FERPA, and that utilized by a privacy technology, differential privacy. This argument is grounded in a legal analysis of FERPA and uses mathematical formalism to address uncertainty in interpreting the legal standard.

Instrumenting the law with a modern scientific understanding of privacy, together with precise mathematical language and models, could help guide the development of modern conceptions of privacy in the law. This approach could also facilitate the development and implementation of new privacy technologies that demonstrably adhere to legal requirements for privacy protection. Introduction of combined legal-technical tools for privacy analysis such as these may also help lawmakers articulate privacy desiderata that more closely match today's information regime, in which a large number of information sources can be accessed and analyzed in ways that lead to breaches of individual privacy. Such techniques can provide robustness and immunity with respect to unknown future privacy attacks, providing long-term, future-proof privacy protection in line with technical, legal, and ethical requirements for protecting the privacy of individuals when handling data about them.

A Proving that differential privacy satisfies FERPA

In Section 4, we extracted a formal model of FERPA's privacy requirements, as well as a mathematical definition of a sufficient level of privacy protection provided by a computation over non-public student data in this model. The model is envisioned as a privacy game, in which a hypothetical adversary tries to beat a computation by using the output of the computation to successfully guess the private information of a specific student whose records were part of the input data fed to the computation.

In Section 4.7.2, we explained that it would be unreasonable to expect an adversary to lose every time, even when playing against a computation that provides "perfect privacy," since the adversary always has some possibility of winning without having access to the output of the computation output—if only by random chance. Instead, we consider a computation to provide privacy protection if no adversary can win the game against it "too often," the intuition being that no adversary should be substantially more successful at guessing student information when she has access to the output of the computation than some adversary would have been without having access to the output. In this discussion, we used the term "real-world scenario" to describe the game in which an adversary plays against a computation while having access to the computation output, and the term "ideal-world scenario" to describe the game in which an adversary plays against a computation *without* having access to the computation output. These scenarios were illustrated in Figures 7 and 8, respectively.

This modeling exercise provided us with a precise, mathematical definition of privacy. By proving that a statistical computation meets the definition of privacy we extracted, we know that it provides a sufficient level of privacy protection to satisfy the privacy requirements of FERPA (as we have modeled them). In this section we show that every differentially-private computation meets this definition of privacy. That is, for every real-world adversary playing the game against a differentially-private computation, there exists an ideal-world adversary (who does not see the computation output) that wins the game with nearly the same probability.

We show this through a sequence of games, which we enumerate H_0 through H_4 and show in Figure 9. We call these games "hybrid" because they represent steps between the ideal-world (H_0, identical to the game in Figure 8) and real-world (H_4, identical to the game in Figure 7) scenarios. Our claim is that, when C is a differentially-private computation, the adversary in the real world does only a little better than an adversary in the ideal world. To substantiate this claim, we consider any pair of consecutive hybrid games, denoted H_i and H_{i+1}, and demonstrate that, for any adversary participating in the hybrid game H_{i+1}, there exists an adversary for the hybrid game H_i with an identical or almost equal winning probability. Intuitively, this means that the difference in privacy loss between two consecutive games is null or very small, and hence it is also the case that the accumulated difference between the privacy loss in the ideal-world game, H_0, and the real-world game, H_4, is small. Given that there is no privacy loss in the ideal-world game, we conclude that the privacy loss in the real-world game is sufficiently small.

We use some mathematical notation in the diagrams and in our discussion below. We use $\bar{d} = (d_1, \ldots, d_n)$ to denote a list of directory information, and d_i to denote the directory information for student i. Similarly, $\bar{P} = (P_1, \ldots, P_n)$ denotes the list of distributions over private student information, and P_i denotes the distribution that describes student i's private information. We use p_i to denote the "actual" value for student i's private information that is sampled from P_i. In addition, s denotes the targeted student, and f is a function over private student information, representing the adversary's goal. DB represents a database of student information. The ith row

in DB consists of d_i and p_i; that is, that ith record consists of the directory information for student i and student i's private information. Finally, c denotes the computation output and g_s is the adversary's guess about the private information of student s. Although the notation is different, H_0 describes the ideal world given in Figure 8, and H_4 describes the real world given in Figure 7.

The argument proceeds by showing that there is little to no privacy loss between the games appearing consecutively in this proof. We review each of the hybrid games in turn.

Hybrid H_0 (the ideal-world game): The starting point of our argument is the the ideal-world game. In this scenario, the game mechanics generate a database with the student information, but perform no computation on it and pass no information on to the adversary. The adversary makes a guess about the private information of the targeted student using only his *a priori* knowledge about the student.

Hybrids H_0 and H_1: The first step is a modification of the ideal-world hybrid H_0. In the ideal-world game, the game mechanics generate the database of all students; this database is referred to as DB. In H_1, the game mechanics also generate a second, related, database $\widehat{DB}$ that is identical to the original DB except that the entry corresponding to the attacked student, s, is removed.

In comparing how an adversary A^* fares in the two games, it is important to note that no information flows from the game mechanics to the adversary in either the ideal-world game or H_1. Intuitively, this implies that the privacy loss in H_1 is no higher than that in H_0. More formally, for any adversary A^* the winning probability is identical in both games.

The only difference between H_0 (the ideal world) and H_1 occurs internally within the game mechanics. In H_0, the game mechanics create a database DB of student information from the directory information $\bar{d}$ and private student information p_i sampled from the distributions over student information $\bar{P} = (P_1, \ldots, P_n)$. The game mechanics in H_1 also take these steps. The game mechanics then create a new database $\widehat{DB}$ that is identical to DB, except that the record of student s has been removed.

All entries of $\widehat{DB}$ are chosen independently from the value p_s of student s. These entries include $d_{s'}, p_{s'}$ for any student s' different than s. Because all of these values are independent of p_s, revealing $\widehat{DB}$ fully or in part would not yield any information about p_s.

Hybrids H_1 and H_2: Note that no information flows from the game mechanics to the adversary in either H_1 or H_2. It follows that for any adversary A^* the winning probability is identical in both games. The only difference between H_2 and H_1 in that in H_2 the computation C is actually run on a database. However, this database is $\widehat{DB}$, which, as noted above, does not carry any information about p_s. It follows that the outcome $\hat{c} = C(\widehat{DB})$ is independent of p_s and revealing $\hat{c}$ would not yield any information about p_s.

Hybrids H_2 and H_3: H_3 differs from H_2 in that the adversary A in H_3 sees the result of C computed on $\widehat{DB}$ before making his guess while the adversary A^* in H_2 does not get any information from the game mechanics. However, by our note above, the outcome of the computation $\hat{c} = C(\widehat{DB})$ does not carry any information about p_s and hence does not help A obtain any advantage over A^*.[262]

[262] Making this argument formally is a bit more subtle. Recall that we need to demonstrate that, for any adversary

Figure 9: Progression of hybrid games.

Hybrids H_3 and H_4: The real-world game differs from H_3 in that, in the real-world game, the game mechanics do not create a database from which the information of student s is removed. Instead, the game mechanics invoke C on the database DB containing the actual data for student s, and give this result to the adversary. As DB contains the student's private information p_s, there is a risk that the output of the computation, which depends on DB, conveys information about p_s, and therefore a risk to the privacy of student s.

To show that this is not the case, we now use the fact that C is a differentially-private computation. Recall from Section 3.1 that if a computation is differentially-private, then its outputs when invoked on two neighboring databases (i.e., databases that differ on one entry) will be similar. In our case, this implies that, for any adversary A, the winning probability in the real-world game is similar to the winning probability in game H_3.

Hybrid H_4 (the real-world game): In this scenario, the computation is performed on the actual student data and the adversary receives the computation output. The adversary can use any knowledge gained by seeing the computation output to more accurately guess the private information of the targeted student. As we have argued, seeing the output of the computation performed on the actual student data does not advantage the adversary very much compared to game H_3 if the computation is differentially-private.

To sum up, we have demonstrated that no privacy is lost transitioning from H_0 to H_3, and that there is only a small loss in privacy stepping from H_3 to H_4. Consequently, the privacy loss in the real world (H_4) is not significantly higher than that in the ideal world (H_0). Since there is no privacy loss in the ideal-world game, it follows that there is no significant privacy loss also in the real-world game. Note that the argument we have given holds for any differentially-private computation; thus, every differentially-private computation meets the privacy definition given in section 4.8.2 and fulfills the privacy requirements of FERPA, up to the limitations of our model.

A participating in H_3, there is an adversary A^* participating in H_2 that has the same winning probability. This is done by the adversary A^* acting identically to A in its communication with the game mechanics. However, because A is expecting to see the outcome c which is not provided to A^*, a simulated $\tilde{c}$ is produced as follows: A^* samples database $\widetilde{DB}$ consisting of the directory information of all students except s and fake entries for these students $\tilde{p}_i$ sampled from P_i. Then, A^* applies C on $\widetilde{DB}$ to obtain $\tilde{c} = C(\widetilde{DB})$ which it passes to A.

B Extensions of the privacy game

In Section 4.12, we explained how the model presented in Section 4.8 can be extended to account for even stronger adversaries. In particular, we contemplated adversaries who can choose which student to target *after* seeing the computation output and who have access to the output of *multiple* computations performed on the student data. In this section, we describe these extensions in further detail and sketch out how they might be modeled.

We note that the privacy definitions for these extensions are more mathematically involved than the definition considered in Section 4.8 and used in Appendix A. Nevertheless, these definitions can be formalized, and an argument analogous to that presented in Appendix A can be made that differential privacy also provides the required protection with respect to these extensions.

B.1 Untargeted adversaries

Recall that Figure 7 above depicts a game we refer to as *targeted*. In this game, the adversary has a particular student s in mind, and the adversary's goal is to improve on guessing this specific student's private information. For instance, the adversary may examine publications of educational data from a Washington D.C. school with the purpose of learning sensitive information about the President's daughter's grades. This *targeted* aspect of the adversary is manifested in the game by the adversary's declaration of the target student s and the target function f. The privacy definition presented in Section 4.8.2 refers to targeted adversaries, and a computation C that satisfies the requirements of the definition provides protection against such adversaries.

We can modify the game in Figure 7 to also consider *untargeted* adversaries, i.e., adversaries that are interested in learning the non-directory personally identifiable information of *any* student in the dataset. An example of such a scenario could be a data broker mining a data release for private information about any student it can. The data broker's goal is to glean non-publicly available information from the data release that it can sell to third parties such as marketers. Although the data broker wants to learn information specific to individual students, the data broker is not committed to targeting a particular student. Instead, it is hoping to learn private information specific to any student in the dataset. Hence, the game and definition presented in Section 4.8—which would require the data broker to commit upfront to targeting a single student—do not capture this scenario.

B.1.1 Mechanics

The untargeted game is shown graphically in Figure 10 below. As before, the adversary chooses a directory of public student information and associates with each student in that directory a probability distribution that describes the adversary's beliefs about the non-directory personally identifiable information of that student. Unlike the previous game, at this point the adversary does not commit to trying to guess the non-directory PII of any particular student. Instead, the adversary passes just the directory and the probability distributions to the game mechanics.

The game mechanics build a database of non-directory student information by sampling randomly from each of the probability distributions given by the adversary and pass this database, with the directory, to a computation C, which outputs some result.

The game mechanics then give this result to the adversary. After seeing this result, the adversary chooses a student to try to attack, and makes a guess about an aspect of that student's non-directory

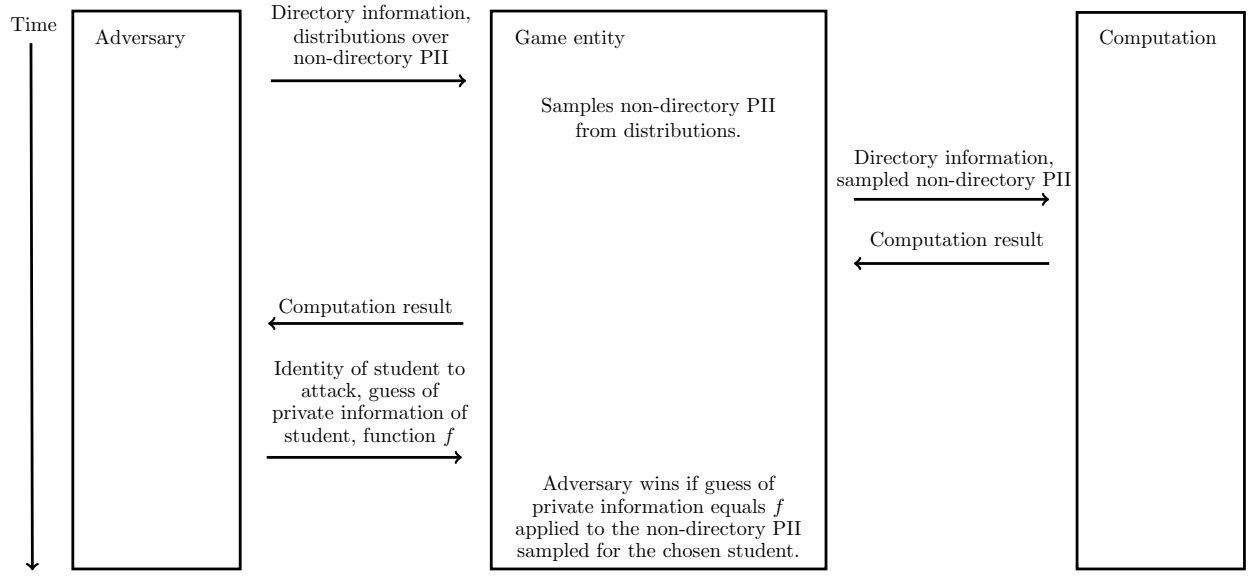

Figure 10: Untargeted scenario. Note that the adversary does not choose a student to attack and a function f until after seeing the computation output.

personally identifiable information. The adversary gives the identity of this student, a function f that it tries to guess, and the guess to the game mechanics. The game mechanics then declares that the adversary has won if his guess matches the result of applying the function f to that chosen student's non-directory personally identifiable information. Otherwise the adversary is considered to have lost the game.[263]

B.1.2 Discussion of the untargeted scenario

The privacy definition for the untargeted scenario is more subtle and mathematically involved than the definition given in Section 4.8.2 for the targeted scenario. We omit the detailed definition and proof from this article.

It is important to note, however, that while protection against untargeted adversaries implies protection against targeted adversaries, it is not the case that protection against targeted adversaries necessarily implies protection against untargeted adversaries. To see why this is the case, consider a hypothetical privacy protection mechanism that chooses a random student among a district's $10,000$ students and publishes this student's entire record. A targeted adversary would only have a chance of one in $10,000$ to learn new information about the targeted student. An untargeted adversary, on the other hand, would always be able to learn new information about some student.[264]

[263] For a more detailed explanation, see *supra* note 243.

[264] We note that the weak privacy protection considered in this section is given only as an example. The authors do not recommend employing such weak protection in practice.

B.2 Multiple releases

The second extension is designed to account for scenarios in which an adversary has access to the outputs of multiple computations on student data. Our formalization currently only accounts for a single release. That is, we assume that the adversary only has access to the output of a single computation performed on the private data. However, it is also important for a robust model to consider the multiple release scenarios. In fact, the FERPA regulations require that, before de-identified data can be released, a determination must be made that "a student's identity is not personally identifiable, whether through single or *multiple* releases."[265]

We can model this requirement with a game in which the adversary sees the output of multiple computations performed on the student data before guessing the private student information. Such a game is given in Figure 11. Like the targeted game presented in Section 4.8, the game starts with the adversary supplying the directory information, distributions over private student information, the identify of a student to attack, and a function over student information f. The game entity creates a database of student information from the directory information and samples drawn from the private attribute distributions. Unlike the previous games we have presented, here the game entity invokes multiple computations (which we enumerate C_1 through C_m) on the database. After each computation, the game entity passes the computation result to the adversary. After all computations have been completed, the adversary makes guesses about the private information of the targeted student and wins if the guess equals f applied to the student's database entry.

This formulation of the game reflects a targeted scenario, where the adversary commits ahead of time to a student to attack. One could alternatively formulate the game for an untargeted scenario with multiple releases, where the adversary only decides which student to attack after he has seen the output of all the computations. Either scenario could be further modified by allowing the adversary to adaptively choose the computations to be performed. This game precedes as follows. The adversary first chooses an initial computation C_1. Thereafter, after seeing each c_i (the result of computation C_i), the adversary chooses C_{i+1}, the next computation to be performed on the student data.

[265] 34 C.F.R. § 99.31(b)(1) (emphasis added).

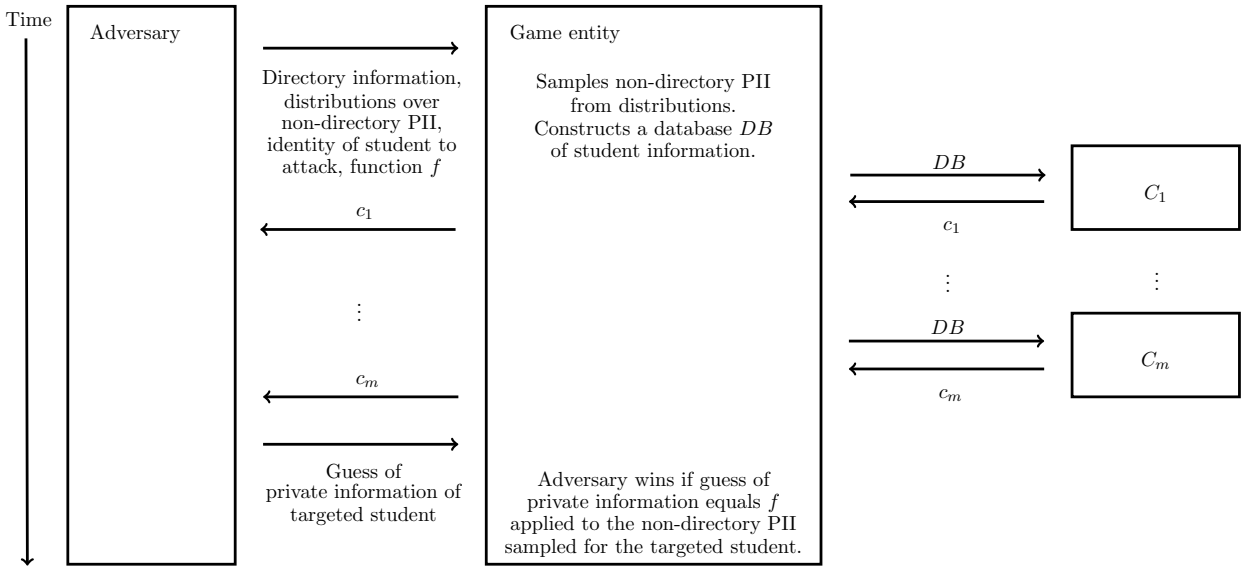

Figure 11: Multiple release scenario.