EPSILON-DIFFERENTIAL PRIVACY, AND A TWO-STEP TEST FOR QUANTIFYING REIDENTIFICATION RISK

Nathan Reitinger[*] and Amol Deshpande[**]

ABSTRACT: Sharing data in the twenty-first century is fraught with error. Most commonly, data is freely accessible, surreptitiously stolen, and easily capitalized in the pursuit of monetary maximization. But when data does find itself shrouded behind the veil of "personally identifiable information" (PII), it becomes nearly sacrosanct, impenetrable without consideration of ambiguous (yet penalty-rich) statutory law—inhibiting utility. Either choice, unnecessarily stifling innovation or indiscriminately pilfering privacy, leaves much to be desired.

This Article proposes a novel, two-step test for creating futureproof, bright-line rules around the sharing of legally protected data. The crux of the test centers on identifying a legal comparator between a particular data sanitization standard—differential privacy: a means of analyzing mechanisms that manipulate, and therefore sanitize, data—and statutory law. Step one identifies a proxy value for reidentification risk which may be easy calculated from an ε-differentially private mechanism: the guess difference. Step two finds a corollary in statutory law: the maximum reidentification risk a statute tolerates when permitting confidential data sharing. If step one is lower than or equal to step two, any output derived using the mechanism may be considered legally shareable; the mechanism itself may be deemed (*statute*, ε)-differentially private.

The two-step test provides clarity to data stewards hosting legally or possibly legally protected data, greasing the wheels in advancements in science and technology by providing an avenue for protected, compliant, and useful data sharing.

CITATION: Nathan Reitinger & Amol Deshpande, *Epsilon-Differential Privacy, and a Two-Step Test for Quantifying Reidentification Risk*, 63 JURIMETRICS J. 263–317 (2023).

[*] Ph.D. Candidate, University of Maryland, Department of Computer Science; M.S., Columbia University; J.D., *magna cum laude*, Michigan State University. This Article benefitted from the Privacy Law Scholars Workshop, 2022. For detailed comments on prior drafts, the authors thank Michael Hawes, Steven Bellovin, Ido Sivan-Sevilla, and Rachel Cummings. The editors of *Jurimetrics* also provided invaluable assistance in the preparation of this Article.

[**] Professor of Computer Science at the University of Maryland at College Park, with a joint appointment in the University of Maryland Institute for Advanced Computer Studies (UMIACS).

One of the most popular data sanitization concepts of the twenty-first century is differential privacy. The idea may be stated in one line: purposefully failing to see the trees for the forest. Differential privacy allows one to learn the statistics of a group without also learning the statistics of the individuals making up the group.[1] To illustrate why this type of crowd-but-not-individual learning is necessary in our data-agnostic[2] world, consider the case of Merck's blockbuster[3] drug Vioxx.[4]

A pharmaceutical boon, Vioxx grossed drug maker Merck around eight billion dollars in the approximately four years it was prescribable, 1999–2004;[5] the drug was marketed as a safer alternative to ibuprofen and quickly became heavily prescribed.[6] All that changed in September 2004 when Merck voluntarily,

[1] Differential privacy operates on the "crowd" level, ensuring that individual data is, in many ways, meaningless. In this way, the individuals who make up the crowd are assured that their data, their secrets, will be kept confidential. *See* Cynthia Dwork & Aaron Roth, *The Algorithmic Foundations of Differential Privacy*, 9 FOUNDS. & TRENDS THEORETICAL COMPUT. SCI. 211, 215–16 (2013).

[2] Formally defined, data agnosticism means that a device is able to receive, as input, many different types of data validly; the device will work regardless of what type of data it is receiving. *See* Tian J. Ma et al., *Big Data Actionable Intelligence Architecture*, 7 J. BIG DATA, 2020, art. no. 103, at 1 (discussing a piece of data-agnostic software capable of accepting several types of incoming data). The saturation of computing devices in our day-to-day life is data-agnostic—all types of data may be validly captured and used to infer attributes about the subjects providing the data. For instance, your smartphone may natively receive cell signal data, allowing you to make phone calls and send text messages, something which may be later used by law enforcement to track your location. However, this type of data requires a warrant, and police cannot request this information without one. *See* Carpenter v. United States, 138 S. Ct. 2206 (2018). Likewise, your smart phone is also collecting data from a large swath of sensors (e.g., gyroscope, magnetometer, and accelerometer to name a few), which enables tracking, behavior detection (e.g., walking, sitting, running), and possibly psychological-type diagnosis like depression. *See* Anupam Das et al., *The Web's Sixth Sense: A Study of Scripts Accessing Smartphone Sensors*, *in* CCS '18: PROCEEDINGS OF THE 2018 ACM SIGSAC CONFERENCE ON COMPUTER AND COMMUNICATIONS SECURITY 1515 (2018); Michalis Diamantaris et al., *The Seven Deadly Sins of the HTML5 WebAPI: A Large-Scale Study on the Risks of Mobile Sensor-Based Attacks*, ACM TRANSACTIONS ON PRIV. & SEC., Nov. 2020, art. no. 19, at 19:1; Taylor A. Braund et al., *Smartphone Sensor Data for Identifying and Monitoring Symptoms of Mood Disorders: A Longitudinal Observational Study*, 9 JMIR MENTAL HEALTH, no. 5, 2022, art. no. e35549, at 1.

[3] *See* James Brumley, *Biggest Blockbuster Drugs of All Time*, KIPLINGER (Jan. 3, 2018), https://www.kiplinger.com/article/investing/t052-c000-s001-biggest-blockbuster-drugs-of-all-time.html [https://perma.cc/XQ2W-FG3M] ("A prescription drug that surpasses $1 billion in sales is known as a blockbuster. It's a rare feat, but when a pharmaceutical company finds a true blockbuster drug the payoff is enormous.").

[4] Richard Knox, *Merck Pulls Arthritis Drug Vioxx from Market*, NPR: ALL THINGS CONSIDERED (Sept. 30, 2004, 12:00 AM), https://www.npr.org/2004/09/30/4054991/merck-pulls-arthritis-drug-vioxx-from-market [https://perma.cc/NTY3-2J4D].

[5] *See* Kurt W. Rotthoff, *Product Liability Litigation: An Issue of Merck and Lawsuits Over Vioxx*, 20 APPLIED FIN. ECON. 1867, 1867–68 (2010) ("When Vioxx was withdrawn, Merck had approximately 9 more years of patent life left on a drug selling $2.5 billion a year"). At the time it was recalled, the drug had been taken by approximately 20 million people. *See* Snigdha Prakash & Vikki Valentine, *Timeline: The Rise and Fall of Vioxx*, NPR (Nov. 10, 2007, 2:40 PM), https://www.npr.org/2007/11/10/5470430/timeline-the-rise-and-fall-of-vioxx [https://perma.cc/WY7B-DB5P] (highlighting events leading up to the recall).

[6] *See* Matthew Herper, *Merck Withdraws Vioxx*, FORBES (Sept. 30, 2004, 8:33 AM), https://www.forbes.com/2004/09/30/cx_mh_0930merck.html?sh=52a1d8c753fb [https://perma.cc/WGQ7-KMP9].

abruptly pulled Vioxx from the market, resulting in a nearly twenty-seven percent stock drop.[7]

While Vioxx had great success at treating arthritis, it also had great success at causing heart attacks.[8] This was a surprise to many—but not the Food and Drug Administration (FDA).[9] Almost three and a half years before Merck imposed its to-be-permanent moratorium on Vioxx sales, the FDA possessed data which, if analyzed, would have illuminated this danger, possibly preventing future harm.[10]

The specter of confidentiality is one of the many reasons why this data never saw the light of day. Releasing legally protected or even potentially legally protected data into the wild is a low-benefit, high risk endeavor. The benefits rely on unpredictable, open-source engagement while the risks involve highly likely public scrutiny and legal backlash (e.g., the mires of anonymization from the 2004s).[11] To be sure, without tools like differential privacy, the risks may outweigh the benefits, and it may be reasonable to withhold data.

Differential privacy solves the confidentiality problem by offering future-proof guarantees to individuals, and therefore data stewards.[12] Differential privacy can guarantee that releasing your data will not be the cause of any adverse

[7] Walter Hamilton, *Merck's Shares Sink 27% on Vioxx News*, L.A. TIMES (Oct. 1, 2004, 4:12 AM), https://www.latimes.com/archives/la-xpm-2004-oct-01-fi-merck1-story.html [perma.cc/NX5G-SC8K].

[8] Daniel J. DeNoon, *When Did Merck Know Vioxx Was Deadly?*, MYTWINTIERS, https://www.mytwintiers.com/news-cat/when-did-merck-know-vioxx-was-deadly-3/ [https://perma.cc/G5Y4-EMEP] (Nov. 30, 2012, 10:17 PM) ("Vioxx increased the risk of [a cardiovascular thromboembolic event] or death by 43%.").

[9] According to the then director for the FDA's center on Drug Evaluation and Research, Dr. Steven Galson, "This [was] not a total surprise." *Merck Recalls Vioxx*, WA. TIMES, Sept. 30, 2004, https://www.washingtontimes.com/news/2004/sep/30/20040930-100336-3743r/ [https://perma.cc/PB2B-YLYC].

[10] *See* Christopher J. Morten & Amy Kapczynski, *The Big Data Regulator, Rebooted: Why and How the FDA Can and Should Disclose Confidential Data on Prescription Drugs and Vaccines*, 109 CALIF. L. REV. 493, 496 (2021) (discussing how evidence of these risks, as identified in the FDA data, became available to experts only after legal proceedings were initiated); *see also id.* at 496 (citing David J Graham et al., *Risk of Acute Myocardial Infarction and Sudden Cardiac Death in Patients Treated with Cyclo-Oxygenase 2 Selective and Non-Selective Non-Steroidal Anti-Inflammatory Drugs: Nested Case-Control Study*, 365 LANCET 475, 480 (2005)); Carolyn Abraham, *Vioxx Took Deadly Toll: Study*, GLOBE & MAIL (Jan. 25, 2005), https://www.theglobeandmail.com/life/vioxx-took-deadly-toll-study/article1113848/ [https://perma.cc/5ZYF-3JJY]; Alex Berenson et al., *Despite Warnings, Drug Giant Took Long Path to Vioxx Recall*, N.Y. TIMES (Nov. 14, 2004), https://www.nytimes.com/2004/11/14/business/despite-warnings-drug-giant-took-long-path-to-vioxx-recall.html [https://perma.cc/BJE2-2SW2] (reporting that over 27,000 heart attacks or deaths were linked to Vioxx).

[11] *See* Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1717–22 (2010) (discussing headliner deidentification affairs like the Netflix prize and AOL search queries).

[12] We use the term *data steward* to refer to any entity which possesses user data. *See generally* Christine L. Borgman, *Open Data, Grey Data, and Stewardship: Universities at the Privacy Frontier*, 33 BERKELEY TECH. L.J. 365, 368–69 (2018) (discussing data stewardship in the context of academic universities).

harm[13]—a guarantee which is not matched by any other sanitization technique.[14] However, while removing the risk in the to-release-or-not-to-release dilemma should result in a positive-sum game, differential privacy today is a mixed bag.

To be sure, differential privacy can be (mostly) life changing[15] *or* something of a dubitante.[16] We see the census going all-in on differential privacy and

[13] Another framing for this may be that adverse effects are quantified, meaning that a data steward can measure the amount of adverse effects to ensure the effects are minimal. *See* Dwork & Roth, *supra* note 1, at 215.

[14] The term *technique* here is used loosely. We are generally referring to ways to think about gaining privacy, while maintaining utility in datasets. More specifically, standards like *k*-anonymity, *l*-diversity, and *t*-closeness provide a way of reasoning about privacy in a sanitized dataset, and are therefore similar to differential privacy, but do not provide the same type of guarantee when it comes adverse harm from an individual's perspective. *See infra* notes 57–62 and accompanying text.

[15] In the technical world, differential privacy is something of a panacea, rarely experiencing critique outside of the constructive flavor, with blind spots shored up by new techniques or extensions that improve the concept over time. *See, e.g.*, Úlfar Erlingsson et al., *RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response*, *in* CCS '14: PROCEEDINGS OF THE 2014 ACM SIGSAC CONFERENCE ON COMPUTER AND COMMUNICATIONS SECURITY 1054, 1054 (2014) (describing a technique for private data collection among end users); Michael Carl Tschantz et al., *SoK: Differential Privacy as a Causal Property*, *in* 2020 IEEE SYMPOSIUM ON SECURITY AND PRIVACY 354, 354 (2020) (concluding that the misgivings about differential privacy are misplaced—issues with trusting differential privacy are actually not issues with differential privacy itself, but instead come from a misunderstanding regarding "correlation doesn't imply causation."); Maritz Hardt & Guy N. Rothblum, *A Multiplicative Weights Mechanism for Privacy-Preserving Data Analysis*, *in* 2010 IEEE 51ST ANNUAL SYMPOSIUM ON FOUNDATIONS OF COMPUTER SCIENCE 61, 61 (2010) ("Our primary contribution is a new differentially private multiplicative weights mechanism for answering a large number of interactive counting (or linear) queries that arrive online and may be adaptively chosen.") (emphasis omitted). *But see* Josep Domingo-Ferrer et al., *The Limits of Differential Privacy (and Its Misuse in Data Release and Machine Learning)*, COMMC'NS. ACM, July 2021, at 33, 33 (finding that gaps in differential privacy are not owed to a fundamental lack of privacy, but instead to a lack of utility degrading its usefulness, suspect assumptions about the correlations among the data, or high values of epsilon used in practice).

[16] On the legal side, however, differential privacy seems to have unwittingly subscribed to a vendetta, where proponents of the tool argue its application in an expansively impressive array of settings, but dissenters appear to be making an attempt to critique the tool out of existence. *Compare* Andrew Chin & Ann Kleinfelter, *Differential Privacy as a Response to the Reidentification Threat*, 90 N.C. L. REV. 1417, 1427–28 (2010) (discussing the "solution" of differential privacy as a response to the "problem" of reidentification); *and* Ohm, *supra* note 11, at 1756 (discussing potential limitations of differential privacy, in a relatively positive light); Felix Wu, *Defining Privacy and Utility in Data Set*s, 84 U. COLO. L. REV 1117, 1137–38 (2013) (discussing differential privacy in a positive light); *and* Anna Myers & Grant Nelson, *Differential Privacy: Raising the Bar*, 1 GEO. L. TECH. REV. 135, 135–36 (2016) ("Differential privacy is raising the bar for effective data responsibility by redefining the balance and reducing the trade-off between privacy and data utility."), *with* Jane Bambauer et al., *Fool's Gold: An Illustrated Critique of Differential Privacy*, 16 VAND. J. ENT. & TECH. L. 701, 701–07, 753–55 (2014) (attempting to curb enthusiasm for differential privacy by highlighting its limitations in regard to practical use cases). And this same dichotomy may be seen through the 2020 census lawsuit. *See* Alabama v. U.S. Dep't Com., No. 3:21-cv-211-RAH-ECM-KCN (M.D. Ala., June 29, 2021); Brief for Jane Bambauer as Amici Curiae Supporting Plaintiffs', Alabama v. U.S. Dep't Comm., 2:21-cv-00211-RAH-ECM-KCN *1, *2 (2021) (arguing for the inapplicability of differential privacy); *see also* Christopher T. Kenny et al., *The Use of Differential Privacy for Census Data and its Impact on Redistricting: The Case of the 2020 U.S. Census*, 7 SCI. ADV., no. 41, 2021, art. no. eabk3283, at 1, 1 ("We find that the [disclosure avoidance system] systematically undercounts the population in mixed-race and mixed-partisan precincts, yielding unpredictable racial and partisan biases."). *See generally* Steven Ruggles & David Van Riper, *The Role of Chance in the Census Bureau Database Reconstruction Experiment*, 41 POPULATION RSCH.

technical scholars having difficulty mentioning privacy without also discussing differential privacy[17]—like a *Marbury v. Madison* for a mathematical understanding of privacy itself.[18] On the other hand, we also see attempts at critiquing differential privacy into the dustbin.[19] Both in academia and in the courts, differential privacy has seen its fair share of challenges. Pundits and politicians aside, there may be some justification for this polarity.

POL'Y REV. 781 (2022). As Jason J. Czarneki notes, "A dubitante (pronounced d[y]oo-bi-tan-tee) opinion indicates that 'the judge doubted a legal point but was unwilling to state that it was wrong.'. . . [T]he judge is unhappy about some aspect of the decision rendered, but cannot quite bring himself to record an open dissent." Jason J. Czarnezki, *The Dubitante Opinion*, 39 AKRON L. REV. 1, 1–2 (2006) (citing BLACK'S LAW DICTIONARY 515 (7th ed. 1999) (emphasis omitted); LON FULLER, ANATOMY OF THE LAW 93 (1968)).

[17] The inventors of the technique (Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith) won the Gödel prize in 2017. *2017 Godel Prize*, EUR. ASS'N THEO. COMP. SCI., https://www.eatcs.org/index.php/component/content/article/1-news/2450-2017-godel-prize [https://perma.cc/CE5Y-KC4Z]; *see* Thomas D. Grant & Damon J. Wischik, *Show Us the Data: Privacy, Explainability, and Why the Law Can't Have Both*, 88 GEO. WASH. L. REV. 1350, 1413 n.228 (2020) ("[Differential privacy] has stood the test of time: in the workshop on privacy in machine learning held as part of the 2019 NeurIPS conference, 25 out of the 42 accepted papers concerned differential privacy."). And most recently, differential privacy is starting to be folded into programming packages to allow for its easy application in software. *See, e.g.*, Naoise Holohan et al., *Diffprivlib: The IBM Differential Privacy Library*, ARXIV 1, 1 (July 4, 2019), https://arxiv.org/pdf/1907.02444.pdf. OpenDP, spawned from Harvard's Privacy Tools Project, also has a suite of open-source differential privacy tools. *About*, OPENDP, https://opendp.org/about [https://perma.cc/D3Y4-8VQK]; Benjamin I.P. Rubinstein & Francesco Aldà, *Diffpriv: An R Package for Easy Differential Privacy*, 18 J. MACHINE LEARNING RSCH., 2017, at 1, 1, https://www.bipr.net/diffpriv/articles/diffpriv.pdf [https://perma.cc/FV3L-BZWD].

[18] Differential privacy, in a way, defines the concept of privacy, similar to how *Marbury v. Madison*, in a way, defines the concept of judicial review. *See* Marbury v. Madison, 5 U.S. 137, 138, 177–80 (1803) (establishing the principle of judicial review). Indeed, it is often difficult, from a technical perspective, to discuss the concept of privacy without bringing up differential privacy. *See, e.g.*, Fang Liu, *A Statistical Overview on Data Privacy*, 34 NOTRE DAME J.L. ETHICS & PUB. POL'Y 477, 479–80 (2020) (introducing privacy leakage by discussing differential privacy's bounded loss). For a discussion of what bounds mean, see *infra* Section II.D.

[19] *See Alabama*, 3:21-cv-211-RAH-ECM-KCN at *4–5 ("Citing the need to counter advancements in computational power and the threat of sophisticated re-identification and reconstruction attacks, the Bureau announced in September 2017 that it would employ a new and more proactive method of disclosure avoidance for the 2020 Census—'differential privacy.' Differential privacy, the Bureau concluded, is the most efficient method by which it can accomplish both of its goals: adequately protecting respondent information while also preserving the utility of census data.") (internal citations omitted).

For one, differential privacy is not a well-understood concept.[20] Math is hard.[21] Additionally, applying differential privacy in a practical, legal setting is in many ways terra incognita.[22] This Article attempts to solve both problems.

The Article begins by teaching differential privacy using a building blocks approach, starting from the most basic[23] and building to a full mathematical def-

[20] This Article adds to, and hopes to build upon, work which similarly attempts to teach differential privacy to a nontechnical audience. *See* Kobbi Nissim et al., *Bridging the Gap Between Computer Science and Legal Approaches to Privacy*, 31 HARV. J.L. & TECH. 687, 689–93 (2017) (introducing differential privacy and applying it to FERPA as a game-based exercise involving, primarily, a mathematical definition of what FERPA requires to release information); Alexandra Wood et al., *Differential Privacy: A Primer for a Non-Technical Audience*, 21 VAND. J. ENT. & TECH. L. 209, 223 (2018) (providing an overview of differential privacy); *see also* Aloni Cohen & Kobbi Nissim, *Towards Formalizing the GDPR's Notion of Singling Out*, 117 PROCS. NAT'L ACAD. SCI. 8344, 8346 (2020) (discussing singling out in technical terms as it may be understood by the GDPR, with a related discussion of differential privacy). Although this set of work takes great efforts to distill differential privacy in basic, understandable terms, there is likely room for improvement in translation without over-precision—that is, starting lower, with basic building blocks, and being more accommodating with the law's natural ambiguity (i.e., maintaining "it depends" while nonetheless providing useful solutions). Most importantly, this Article attempts to provide something that every One L learns the value of around exam time; understanding a concept in theory is great, but having an attack plan for what to do with a fact pattern is golden. *See* SCOTT TUROW, ONE L: THE TURBULENT TRUE STORY OF A FIRST YEAR AT HARVARD LAW SCHOOL 1 (2010); RICHARD MICHAEL FISCHL & JEREMY PAUL, GETTING TO MAYBE: HOW TO EXCEL ON LAW SCHOOL EXAMS 1 (1999). In this light, this Article has two primary goals: (1) introduce differential privacy in an accessible way; and (2) use that understanding to form the basis of an easy-to-apply, two-step test for attacking legal problems.

[21] Similar to how odd the following statement is at first glance—in room of 23 people there is a 50% chance that two of them have the same birthday—differential privacy only offers an understanding of its protections through a mathematical conversation. *See Understanding the Birthday Paradox*, BETTER EXPLAINED, https://betterexplained.com/articles/understanding-the-birthday-paradox/ [https://perma.cc/7UK7-GL77]; Russell Samora, *The Birthday Paradox*, PUDDING, https://scout.wisc.edu/archives/index.php?P=GoTo&ID=49730&MF=4 (last visited Nov. 7, 2022); *see also* Jackson v. Pollion, 733 F.3d 786, 790 (7th Cir. Ill. 2013) ("To determine the effect on the plaintiff's health of a temporary interruption in his medication, the lawyers in the first instance, and if they did their job the judges in the second instance, would have had to make some investment in learning about the condition. . . . The legal profession must get over its fear and loathing of science. . . . [T]his plainly meritless suit was filed on September 2, 2009—more than four years ago. The intervening years have been consumed largely by procedural wrangling and protracted, tedious depositions. A stronger judicial hand on the tiller could have saved a good deal of time, effort, and paper."). And more directly related to differential privacy, the controversial *Fool's Gold* Article had a particularly galvanizing effect on some. *See* Frank McSherry, *Differential Privacy for Dummies*, GITHUB (Jan. 4, 2017), https://github.com/frankmcsherry/blog/blob/master/posts/2016-02-03.md [https://perma.cc/EYT4-N2JR] ("[The *Fool's Gold*] article starts 'Legal scholars champion differential privacy as a practical solution to the competing interests in research and confidentiality, and policymakers are poised to adopt it as the gold standard for data privacy. It would be a disastrous mistake.' And concludes '[d]ifferential privacy faces a hard choice. It must either recede into the ash heap of theory, or surrender its claim to uniqueness and supremacy.' I will present a third option: The authors could take a fucking stats class and stop intentionally misleading their readers." (citing Bambauer et al., *supra* note 16, at 701, 753–54)).

[22] True enough, there have been several attempts at applying differential privacy in a legal setting. *See supra* note 20 and accompanying text. While attractive in theory, however, these approaches may be difficult to directly apply in a practical, legal setting.

[23] *See infra* Part II (teaching differential privacy using a ground-up approach).

inition.[24] The Article then proposes a novel means of understanding differential privacy, which intends to be readily applicable in a statutory framework.[25] By default, differential privacy does not directly translate to statutes regulating data. If data protection laws made statements like "it is permissible to share data if the mechanism of release applies differential privacy with an epsilon value of less than or equal to .05,"[26] then this Article would be superfluous. For good reason,[27] this will likely never happen. Instead, most statutes create ambiguous mandates like "remove any information which could lead to identification."[28] This, however, leaves a data steward in a difficult position: How much sanitization does a dataset[29] need to undergo before there is no data remaining that could "lead to identification?"[30] Likewise, this leaves differential privacy in a difficult position: When does a differentially private mechanism permit legally shareable data and when does it not?

This translation problem stems equally from issues rooted in both law and technology. To solve it requires finding a common element among data protective statutes that provides a metric against which differential privacy can be measured. Stated otherwise, is there a single, mathematical value that (step one) may be derived from a differentially private mechanism and (step two) is translatable to what statutes require for the sharing of confidential data? Yes.

For the legal piece, all data protective statutes, we argue, regulate "reidentification risk."[31] Statutes go about this by using a variety of unique textual phrases (e.g., "personally identifiable information" (PII) or "personal data") but what all of these phrases have in common is an intent to reduce the potential for

[24] *See infra* Part III (unveiling the full mathematical definition of differential privacy and introducing step one of our two-part test).

[25] *See infra* Section III.C (identifying a proxy value which summarizes the reidentification risk a differentially private mechanism encumbers—the guess difference).

[26] At least one state statute in the United States has been built with differential privacy in mind, but without mentioning specific requirements. *See, e.g.*, WASH. REV. CODE § 69.51A.230(8)(c) (2022) (requiring a medical cannabis database to "incorporate current best differential privacy practices, allowing for maximum accuracy of database queries while minimizing the chances of identifying the personally identifiable information included therein").

[27] Determining the appropriate balancing of interests in the making of law does not go hand in hand with definitive, one-size-fits-all rules. *See* Jarrod Shobe, *Intertemporal Statutory Interpretation and the Evolution of Legislative Drafting*, 114 COLUM. L. REV. 807, 866–67 (2014) (discussing how statutory ambiguity is, today, most often intentional, and plays a role in the rule-making process); *see also infra* note 208 and accompanying text.

[28] *See infra* Section IV.A.1.

[29] A dataset is simply a collection of data, like a series of rows in a spreadsheet. *See infra* note 59 and accompanying text.

[30] Unfortunately, the easiest answer here is simply not asking the question and refusing to share data entirely. *See* David Deming, *Balancing Privacy with Data Sharing for the Public Good*, N.Y. TIMES, (Feb. 19, 2021), https://www.nytimes.com/2021/02/19/business/privacy-open-data-public.html ("Governments and technology companies are increasingly collecting vast amounts of personal data, prompting new laws, myriad investigations and calls for stricter regulation to protect individual privacy. Yet despite these issues, economics tells us that society needs more data sharing rather than less, because the benefits of publicly available data often outweigh the costs.").

[31] *See infra* Part IV (assessing differential privacy from a practical legal perspective and introducing step two: the maximum reidentification risk a statute permits).

harm that an individual described in a dataset faces when their data is shared.[32] The harm is that someone can look at a statutorily compliant dataset and say "this record is your record." Some statutes may have a very low threshold for risk (e.g., Health Insurance Portability and Accountability Act (HIPAA)[33]) while others may have a high tolerance (e.g., Video Privacy Protection Act (VPPA)[34]); ultimately, however, all statutes share a common goal of reducing this risk—at varying thresholds.

For the technical piece, the Article looks at identifying a value native to differential privacy which may be called the reidentification risk of the mechanism. We call this value the guess difference[35]: the risk of being reidentified in data that comes from an ε-differentially private mechanism. This proxy value for reidentification risk is easy to understand, adds context to an otherwise ambiguous number, and allows differential privacy to be directly compared to what statutes mandate in terms of data confidentiality.[36]

Working together, the two-step test provides much needed confidence to data stewards hosting legally protected data. The test permits easy line drawing around how little or how much sanitization is required before sharing data within a regulatory ecosystem—greasing the wheels on private, useful data sharing.[37] Before introducing the primitives on which differential privacy operates, the Article first elaborates on how modern-day privacy leads to, and necessitates, differential privacy.[38]

[32] *See infra* Section IV.A.

[33] Health Insurance Portability and Accountability Act (HIPAA) of 1996, 42 U.S.C. § 1320 (2017); 45 C.F.R. § 164.514 (2021) (requiring, in a safe harbor provision, the stripping of numerous identifiers before the legal release of data).

[34] Video Privacy Protection Act of 1988, 18 U.S.C. § 2710 (2018). VPPA has been deemed by the courts to *not* prevent data release even if some amount of detective work could reidentify an individual. *See* Eichenberger v. ESPN, Inc., 876 F.3d 979, 985 (9th Cir. 2017) ("[I]t was clear that, if the disclosure were that 'a local high school teacher' had rented a particular movie, the manager would not have violated the statute. That was so even if one recipient of the information happened to be a resourceful private investigator who could, with great effort, figure out which of the hundreds of teachers had rented the video.").

[35] For reasons we further elaborate below, we do not wish to call this value "reidentification risk" directly. *See infra* Section III.B. First, differential privacy does not natively provide a "reidentification risk" value; we are attempting to find a proxy value in differential privacy which represents reidentification risk. Second, there may be many ways to express the reidentification risk of a differentially private mechanism. The "guess difference" is merely one of the ways we felt did a particularly good job at being both insightful and simple. We urge further technical work to build upon our Article by analyzing the "guess difference" and possibly coming up with a more robust way of summarizing the reidentification risk a differentially private mechanism provides.

[36] *See infra* Section III.B.2 (discussing the options for guess difference and how some of these options lack context that allows reidentification risk to be understood), Section IV.B (providing an example application of our two-step test).

[37] *See infra* Section IV.C (highlighting the benefits this type of test offers).

[38] *See infra* Part I (explaining why differential privacy is a necessary tool when sharing private data).

I. CONSTITUTIONAL HERITAGE

Data is inescapable,[39] revolutionary,[40] and commoditized.[41] Everything you do online[42] and offline[43] is captured. True enough, this has negative side ef-

[39] *See* Eben Moglen, *Law in the Internet Society*, COLUM. L. (Nov. 03, 2020), https://moglen.law.columbia.edu/audio/LIS-2020-11-03 [https://perma.cc/5P22-4JFE] ("[A]s a result of changes brought about by the cold war . . . in the 1970s, the American government and the research institutions of American society began to evolve the nervous system of a species-wide interconnection for the human race that at the time was thought of only as a method for securing robust political and military command and control under the conditions of possible nuclear war. . . . On top of this physiology is the economy of everything that came to be digital, that is, that came to have zero marginal cost of copying and transmission, which turned out to be pretty much all the cultural possession of the human race, its science, its art, its music, its exchange information, its authenticated records, its history, its journalism, its spying propaganda . . . . This structure of a system of interconnected neural operations that spans the globe and can reach anything has no defined social purpose. It can be used to allow every brain on earth to learn anything . . . it wants . . . . regardless of the ability to pay or govern attitude. It can also be used to perfect surveillance based despotism . . . . Above the layer of the physiology of the switches, the network can be thought of as a hierarchy of services. You get from somewhere the services that through the handset or laptop or other form of switch closest to your eyeball and your eardrum supports your life. Email and other forms of messaging between individuals, calendaring, and the other tools of collaboration which are ethnomethodologies now, ways of living just as important as how you walk down the street without bumping into people or how you use a knife and fork so as to assume the role in one or another society of civilized person. . . . At the center of this network now barely half a decade old [is] this entity we created that knows about human beings . . . that eats behavior in order to increase its grasp, and secrets hormones that create more clicking, swiping, tapping, buzzing, beeping, walking, running and etcetera—this parasite with the mind of God.") (audio recording of class).

[40] *See* Peter F. Drucker, *Beyond the Information Revolution*, ATLANTIC (Oct. 1999), https://www.theatlantic.com/magazine/archive/1999/10/beyond-the-information-revolution/304658/ [https://perma.cc/T3YQ-Y72J]; Yaameen Choudhury, *Data Science: 7 Reasons Why It Is the Most Revolutionary Sector of the Century*, TECHIE CUB (Sept. 7, 2021), https://techiecub.com/data-science-7-reasons-why-it-is-the-most-revolutionary-sector-of-the-century/ [https://perma.cc/Z464-BW83]; Steve MacFeely, *The Data Revolution is Only Beginning*, UNITED NATIONS: UNITE (Aug. 13, 2021), https://unite.un.org/blog/data-revolution-only-beginning [https://perma.cc/J35Y-7PHE]; Andrew Brust, *Why Is Big Data Revolutionary*, ZDNET (Apr. 10, 2012), https://www.zdnet.com/article/why-is-big-data-revolutionary/ [https://perma.cc/DS8S-7EVV].

[41] In 2017, the Economist stated that "the world's most valuable resource is no longer oil, but data." *The World's Most Valuable Resource Is No Longer Oil, but Data*, ECONOMIST (May 6, 2017), http://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource [https://perma.cc/2K7Z-3LH3]; *see also* Dennis D. Hirsch, *The Glass House Effect; Big Data, the New Oil, and the Power of Analogy*, 66 ME. L. REV. 373, 373–75 (2014) ("Data is an essential resource that powers the information economy much like oil has fueled the industrial economy. Big Data promises a plethora of new uses—the identification and prevention of pandemics, the emergence of new businesses and business sectors, the improvement of health care quality and efficiency, and enhanced protection of the environment, to name but a few—just as oil has generated useful plastics, petro-chemicals, lubricants, and gasoline. Big Data 'is becoming a significant corporate asset, a vital economic input, and the foundation of new business models. It is the oil of the information economy.' This Article looks at the analogy in a different way, one not yet developed in the scholarly literature. It examines the underside of the 'Big Data is the new oil' comparison. Oil certainly has many productive uses, but it also leads to oil pollution. Big Data is similar. It produces tremendous benefits, but simultaneously generates significant privacy injuries." (quoting VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK AND THINK 16 (2013)).

[42] *See* Moglen, *supra* note 39; *see also* Nathan Reitinger & Michelle L. Mazurek, *ML-CB: Machine Learning Canvas Block*, 2021 PROC. ON PRIV. ENHANCING TECHS., no. 3, 2021, at 453, 459 (finding that nearly a quarter of the top 100 websites on Alexa Top Rank engaged in canvas

fects—for example, Panopticon-styled chilling effects, overbroad NSA dragnets,[44] and the transactional cost of reductionism in the pursuit of category-themed, ad-based[45] monetization. But at the same time, big data[46] also has positive side effects—for example, democratized education via massive open online courses,[47] the proliferation of e-commerce, and worldwide, instantaneous communication networks. To be sure, no effect (positive or negative) is without a privacy loss.

A. Privacy Loss: Legal Protections

Privacy loss is a difficult-to-describe harm, but one which, when looking for it, may be easily found in marking the boundary lines of governmental intrusion. Marriages, procreation, and parenthood have all been subject to fierce protection (or at the very least debate) by the Supreme Court,[48] which has ex-

fingerprinting, a technique used to *surreptitiously* track website visitors), https://petsymposium.org/popets/2021/popets-2021-0056.pdf [https://perma.cc/7VUT-HM2M]; Chris Jay Hoofnagle et al., *Behavioral Advertising: The Offer You Cannot Refuse*, 6 HARV. L. & POL'Y REV. 273, 273–79 (2012).

[43] *See* Charles Duhigg, *How Companies Learn Your Secrets*, N.Y. TIMES (Feb. 16, 2012), https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html [https://perma.cc/8ZWG-VKCT] (tracking pregnant women based on products purchased rather than online activity); Miranda Wei et al., *What Twitter Knows: Characterizing Ad Targeting Practices, User Perceptions, and Ad Explanations Through Users' Own Twitter Data*, *in* PROCEEDINGS OF THE 29TH USENIX SECURITY SYMPOSIUM 145, 153 tbl.1 (Aug. 2020), https://www.usenix.org/system/files/sec20-wei.pdf (reporting, in Table 1, that some sources of advertising come not from internet-based behaviors like web-browsing history, but from real-world attributes like household income or location).

[44] *See* Megan Pugh, *Note, Privacy? What Privacy?: Reforming the State Secrets Privilege to Protect Individual Privacy Rights From Expansive Government Surveillance*, 9 BELMONT L. REV. 265, 272–73 (2021) ("PRISM is an NSA internet surveillance tool created to collect the private internet data of foreign nationals. However, in doing so, it also sweeps up the data of United States citizens, including emails, files and photos, through accessing user accounts on Gmail, Facebook, Apple, Microsoft and other technology companies."); Mark M. Jaycox, *No Oversight, No Limits, No Worries: A Primer on Presidential Spying and Executive Order 12,333*, 12 HARV. NAT'L SEC. J. 58, 61 (2021).

[45] *See* Ben Weinshel et al., *Oh, the Places You've Been! User Reactions to Longitudinal Transparency About Third-Party Web Tracking and Inferencing*, *in* CCS '19: PROCEEDINGS OF THE 2019 ACM SIGSAC CONFERENCE ON COMPUTER AND COMMUNICATIONS SECURITY 149 (Nov. 2019) (illuminating online tracking mechanisms with a transparency-inspired, tracker-themed browser extension), https://dl.acm.org/doi/pdf/10.1145/3319535.3363200; Wei et al., *supra* note 43 (exploring the ways Twitter views users from a sales perspective—i.e., of the then 30 ways to assess users, advertisers may target users by household income, keywords like "#AfricanAmerican" or "unemployment," or any type of list (with any type of name), such as "Christian Audience to Exclude" or "LGBT Suppression List").

[46] This quotation is from 2017: "The world produces 2.5 quintillion bytes a day, and 90% of all data has been produced in just the last two years." Vasudha Thirani & Arvind Gupta, *The Value of Data*, ECON. F. (Apr. 22, 2017), https://www.weforum.org/agenda/2017/09/the-value-of-data/ [https://perma.cc/UKF2-HDAM]. It is predicted that by 2025 "the amount of data generated each day is expected to reach 463 exabytes globally." *How Much Data Is Created Every Day?*, SEED SCIENTIFIC (Oct. 28, 2021), https://seedscientific.com/how-much-data-is-created-every-day/ [https://perma.cc/6YK4-5GFB].

[47] *See, e.g.*, Meltem Huri Baturay, *An Overview of the World of MOOCs*, 174 PROCEDIA-SOC. & BEHAV. SCIS. 427, 427 (2015).

[48] *See* Griswold v. Connecticut, 381 U.S. 479, 480 (1965) (striking down a Connecticut statute which prohibited the use of contraception); Eisenstadt v. Baird, 405 U.S. 438, 443 (1971) ("In-

plicitly noted the right to privacy as one of the most valued rights for all citizens. For example, the Court stated in *Lawrence v. Texas*: "[There is] *no* legitimate state interest which can justify its intrusion into the individual's personal and private life."[49] And the Court observed in *Stanley v. Georgia*:

> If the First Amendment means anything, it means that a State has no business telling a man, sitting alone in his own house, what books he may read or what films he may watch. Our whole constitutional heritage rebels at the thought of giving government the power to control men's minds.[50]

Though not explicitly tied to the polarizing invasions highlighted in *Lawrence v. Texas* and *Stanley v. Georgia*, data itself makes these types of invasions

stead, the court concluded that the statutory goal was to limit contraception in and of itself—a purpose that the court held conflicted 'with fundamental human rights' under *Griswold v. Connecticut*, 381 U.S. 479 (1965), where this Court struck down Connecticut's prohibition against the use of contraceptives as an unconstitutional infringement of the right of marital privacy."); Roe v. Wade, 410 U.S. 113 (1973) (protecting a women's right to an abortion), *overruled by* Dobbs v. Jackson Women's Health Org., 2022 U.S. LEXIS 3057 (June 24, 2022); Barton Gellman & Sam Adler-Bell, *The Disparate Impact of Surveillance*, CENTURY FOUND. (Dec. 21, 2017), https://tcf.org/content/report/disparate-impact-surveillance/?agreed=1 [https://perma.cc/QQ5C-GD4V] ("The concept of privacy was conceived in counterpoint to the government's growing ambition to peer inside long-closed compartments of our personal lives. The struggle took place most often in court. The leading cases mark milestones in that history, and show the disproportionate place of minority surveillance in the evolution of law.").

[49] Lawrence v. Texas, 539 U.S. 558, 578 (2003) (emphasis added); *see also* DALE CARPENTER, FLAGRANT CONDUCT: THE STORY OF *LAWRENCE V. TEXAS* 3–18 (2013). Perhaps shocking today, these types of laws are anything but uncommon to history:

> Through the nineteenth century and well into the twentieth, every state in the United States had laws prohibiting anal sex, often called in state statutes "crimes against nature," "sodomy," or "buggery." During the same period, states also began specifically prohibiting oral sex. Prior to the late 1960s, such laws applied regardless of the sex of the participants in the act and regardless of whether the couple was married. A husband and wife who engaged in oral sex were potentially as guilty as two men who had anal sex. This reflected the moral view that all sex outside of marriage, and all nonprocreative sex within marriage, were improper expressions of human sexuality.

Id. at 4–5.

[50] *See* Stanley v. Georgia, 394 U.S. 557, 565 (1969).

possible.[51] Data permits oppression[52] just as easily as it permits freedom. If not properly harnessed "data availability" can become "the database of ruin," transforming worldwide communication networks into worldwide surveillance networks.[53] The bright side is that data sanitization has grown in leaps and bounds since the early days of reidentification awareness.[54]

[51] Basic properties of data, like persistence, make it amenable to overreaching. For example, in *Carpenter*, a defendant was tied to the scene of a crime by using fine-grained motion data—data which is necessarily, natively produced when using a cellular phone. Carpenter v. United States, 138 S. Ct. 2206, 2212–13 (2018) ("Altogether the Government obtained 12,898 location points cataloging Carpenter's movements—an average of 101 data points per day."); *see* Rashida Richardson et al., *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*, 94 N.Y.U. L. REV. 15, 19–20 (2019) (assessing "bias in, bias out" in the data used by predictive policing tools); Cameron Martin, *Facial Recognition in Law Enforcement*, 19 SEATTLE J. FOR SOC. JUST. 309, 309 (2020) ("[In] October 2018, Thomas saw a neighbor preparing to cut a tree from Thomas's land. The two of them got into a heated argument . . . . Someone called the police, and a patrolman of the Carrolton Township Police Department (CTPD), Jack Vincennes, responded to the call. Patrolman Vincennes broke up the fight, but the facial-recognition feature of his CopperFR body camera attached to the front of his uniform flagged Thomas. This camera automatically took a picture of Thomas and sent it back to the precinct. The system identified Thomas as another person, Rollo Smith, who was the subject of an outstanding arrest warrant for robbery and murder in Los Angeles, California . . . . Vincennes arrested Thomas, who was held for three days while the Carrolton and Los Angeles Police Departments (LAPD) conducted further investigations. On the fourth day, a comparison of his fingerprints and physical description with the LAPD's records definitively showed that Thomas was not Smith, who, aside from having different fingerprints, also had several distinctive scars and tattoos. Thomas was released."); *see also* Riley v. California, 573 U.S. 373, 393–94 (2014).

[52] *See* Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 673 (2016) ("Where data is used predictively to assist decision making, it can affect the fortunes of whole classes of people in consistently unfavorable ways. Sorting and selecting for the best or most profitable candidates means generating a model with winners and losers. If data miners are not careful, the process can result in disproportionately adverse outcomes concentrated within historically disadvantaged groups in ways that look a lot like discrimination.").

[53] *See* Ohm, *supra* note 11, at 1746. Data is simply a tool wielded to enforce certain goals, appropriate or inappropriate. For example, the *Doomsday Book* was simply a physical database of tax information, considered an oppressive tool of the time given the way it was used to control ownership rights. John Henry Clippinger, *Digital Innovation in Governance: New Rules for Sharing and Protecting Private Information*, *in* RULES FOR GROWTH: PROMOTING INNOVATION AND GROWTH THROUGH LEGAL REFORM 381, 387 (2011) ("The first inkling of Western privacy awareness manifested itself nearly one thousand years ago with the issuance of the *Doomsday Book* (so named after the Anglo-Saxon term 'doom,' for reckoning, accounting, judgment) by the Norman king, William the Conqueror, in 1086. For the first time in the West, a ruler had a written record in Latin of the major property holdings of his subjects. For non-Normans, it was a greatly feared and resented registry, because it gave the Norman king unprecedented powers to tax properties and assemble armies."); *see also* Alistair, *Big Data Is Our Generation's Civil Rights Issue, and We Don't Know It*, SOLVE FOR INTERESTING (July 31, 2012, 12:40 PM), http://solveforinteresting.com/big-data-is-our-generations-civil-rights-issue-and-we-dont-know-it/ [https://perma.cc/DC32-G93B] ("OK cupid's 2010 blog post 'The Real Stuff White People Like' showed just how easily we can use information to guess at race. . . . They simply looked at the words one group used which others didn't often use. The result was a list of 'trigger' words for a particular race or gender. Now run this backwards. If I know you like these things, or see you mention them in blog posts, on Facebook, or in tweets, then there's a good chance I know your gender and your race, and maybe even your religion and your sexual orientation. And that I can personalize my marketing efforts towards you.").

[54] This would be the idea that simple methods of deidentification are not bullet proof, and may nonetheless enable attackers to reidentify individuals in a deidentified dataset. *See* Jules Polonetsky et al., *Shades of Gray: Seeing the Full Spectrum of Practical Data De-Identification*, 56

B. Privacy Loss: Technical Protections

Techniques, tools, and a mountain of interest have matured around the "safe" sharing of data, vis-à-vis standards like *k*-anonymity,[55] *l*-diversity,[56] and *t*-closeness.[57] At a high level, these concepts are all attempts at assessing the privacy preserving qualities of an "anonymized"[58] dataset.[59] In simple terms, taking a set of raw data, applying some measure of noise (e.g., "suppression" by redacting all zip code digits), and then assessing how privacy preserving the resulting dataset is.[60]

The thorn for each of these standards, however, is that *none* of them provide guarantees in the same way that differential privacy provides guarantees.[61] For

SANTA CLARA L. REV. 593, 594–95 (2016) ("Computer scientists and mathematicians have come up with a re-identification tit for every de-identification tat.").

[55] As discussed below, differential privacy provides a way to reason about mechanisms or algorithms which perturb data. *See infra* Part II. Likewise, *k*-anonymity provides a similar viewpoint—how privacy preserving is this dataset if it meets the *k*-anonymity standard, if $k - 1$ records in the dataset are identical? Latanya Sweeney, k-*Anonymity: A Model for Protecting Privacy*, 10 INT'L J. ON UNCERTAINTY, FUZZINESS & KNOWLEDGE-BASED SYS. 557, 559 (2002) ("[The joint-based privacy attack] provides a demonstration of re-identification by directly linking (or 'matching') on shared attributes. The work presented in this paper shows that altering the released information to map to many possible people, thereby making the linking ambiguous, can thwart this kind of attack. The greater the number of candidates provided, the more ambiguous the linking, and therefore, the more anonymous the data.").

[56] *See* Ashwin Machanavajjhala et al., *l-Diversity: Privacy Beyond k-Anonymity*, ACM TRANSACTIONS KNOWLEDGE DISCOVERY DATA, Mar. 2007, art. 3, at 1, 3, https://dl.acm.org/doi/pdf/10.1145/1217299.1217302 ("[D]oes *k*-anonymity really guarantee privacy? In the next section, we will show that the answer to this question is interestingly no. We give examples of two simple yet subtle attacks on a k-anonymous dataset that allow an attacker to identify individual records. Defending against these attacks requires a stronger notion of privacy that we call *l*-diversity, the focus of this article.").

[57] *See* Ninghui Li et al., *t-Closeness: Privacy Beyond k-Anonymity and l-Diversity*, *in* 2007 IEEE 23RD INTERNATIONAL CONFERENCE ON DATA ENGINEERING 106 (Apr. 2006) (improving *l*-diversity by focusing on what information an attacker may have as background knowledge). Newer techniques also exist, such as machine-learning based synthetic data generation. *See* Steven M. Bellovin, Preetam K. Dutta, & Nathan Reitinger, *Privacy and Synthetic Datasets*, 22 STAN. TECH. L. REV. 1, 5 (2019) ("In essence, take an original (and thus sensitive) dataset, use it to train a machine-learning enabled generative model, and then use that model to produce realistic, yet artificial data that nevertheless has the same statistical properties.") (footnotes omitted).

[58] "Anonymization" would in some ways be a misnomer because these techniques do not always provide truly private data. *See* Wu, *supra* note 16, at 1126 ("One cannot talk about the success or failure of anonymization in the abstract. Anonymization encompasses a set of technical tools that are effective for some purposes, but not others. What matters is how well those purposes match the law and policy goals society wants to achieve. That is a question of social choice, not mathematics.").

[59] A dataset (different from a database) may be thought of as a collection of data organized in a tabular (i.e., columns and rows) format. *See* Bellovin et al., *supra* note 57, at 10.

[60] At a high level, this is the same process differential privacy. *See generally* Liu, *supra* note 18, at 477 (discussing sanitization practices).

[61] Although these techniques do provide what may be described as types of guarantees, the guarantees are not very useful in a practical sense. *See, e.g.*, Aaron Beach et al., *Social-K: Real-Time k-Anonymity Guarantees for Social Network Applications*, *in* 8th IEEE INTERNATIONAL CONFERENCE ON PERVASIVE COMPUTING AND COMMUNICATIONS 600 (2010) (discussing the guarantee of *k*-anonymity, that a record is only unique in terms of the set of indistinguishable records it must be grouped with—a guarantee which says little about what type of external attacks may be

example, consider a dataset that is k-anonymized.[62] In shorthand, what this means is that, assuming a k value of three, for every record (i.e., row in a spreadsheet) at least two other records are identical.[63]

If your data were within this anonymized dataset,[64] would you feel comfortable allowing it to be released publicly, into the wild, forever? Your answer likely depends on the sensitivity of the data, the trustworthiness of the data steward, how the data may be used by others, and the many other potential implications arising from releasing the data.[65] Underneath each of these concerns, however, is a singular risk: How likely is it that you will be "reidentified?" In a worst-case scenario, what is the risk that someone will be able to point at your record and say "this is you." All other adverse effects stem from this singular, null-privacy end result: reidentification.[66]

Most sanitization standards, like k-anonymity, say nothing about how likely or unlikely the threat of reidentification is.[67] Is a k value of three, four, or five required for a release to be privacy preserving enough to make the risk of reidentification minimal? If "joins"[68] with auxiliary data are off the table (i.e., a

successful despite the use of a particular k value). For more detail on k-anonymity, see *infra* Section II.A. Even newer methods of sanitized data generation, like vanilla synthetic data, do not provide the type of guarantees that differential privacy provides. *See* Bellovin et al., *supra* note 57, at 37–41 ("[W] would be remiss if we did not make it absolutely clear that synthetic data and even differentially private synthetic data are not silver bullets"); *see also* Theresa Stadler et al., *Synthetic Data—Anonymisation Groundhog Day*, *in* PROCEEDINGS OF THE 31ST USENIX SECURITY SYMPOSIUM 1451 (2022), https://www.usenix.org/system/files/sec22-stadler.pdf (discussing how synthetic data is rarely a silver bullet).

[62] *See* Latanya Sweeney, *Achieving k-Anonymity Privacy Protection Using Generalization and Suppression*, 10 INT'L J. UNCERTAINTY, FUZZINESS & KNOWLEDGE-BASED SYSTEMS 571, 571–72 (2002) ("One way to achieve this is to have the released records adhere to k-anonymity, which means each released record has at least $(k-1)$ other records in the release whose values are indistinct over those fields that appear in external data. So, k-anonymity provides privacy protection by guaranteeing that each released record will relate to at least k individuals even if the records are directly linked to external information.").

[63] More accurately, the identical records are identical in regard to quasi-identifiers (i.e., the attributes that may be used to identify an individual, either by themselves or in combination). For more detail on k-anonymity and a fuller explanation for how the standard works, see *infra* Section II.A.

[64] For example, if a table had columns of name, zip code, date of birth, and 'is or is not' a parent, then the information may look like this in raw form <Nathan, 20009, 06-30-1999, no> and this in k-anonymized form <Nathan, 2****, 06-**-1999, no>. Assuming there are at least two other people named Nathan who have a zip code starting in 2, a birthday in June of 1999, and who are not parents, the dataset is valid, at least for this record, at a k level of 3.

[65] *See generally* CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION 141–60 (2016).

[66] Notably, it is patently true that many other forms of harm exist following the deidentification of some amount of data found in a "sanitized" dataset. For the purposes of *statutory* data sanitization requirements, however, we limit our scope, and therefore our test, to the harms of reidentification exclusively. *See* discussion *infra* Section IV.A.1. Likewise, it is also noteworthy that data protective statutes do much more than regulate the sharing of data. *See, e.g.*, 16 C.F.R. § 312.1 (2020) (predominantly regulating data collection).

[67] *See infra* Section II.A (discussing reidentification in regard to a dataset anonymized with k-anonymity).

[68] This is a specific type of reidentification attack on a dataset which seeks to take unknown information and match it with known information, in the end revealing something about the unknown information. In more technical terms, a join may be thought of as a combination of two

privacy attack which matches known information with unknown information—one which *k*-anonymity was designed to protect against), is the threat of reidentification completely eliminated?[69] Are contractual requirements needed to add teeth to a sanitization technique's gums or are the technical mechanics of sanitization a sufficient deterrent? No standard mentioned above provides definitive answers to these questions *except* for differential privacy.

Differential privacy is no tourist when it comes to guarantees. In fact, differential privacy was built with these guarantees in mind, requiring that individual data, by itself, be meaningless. As Cynthia Dwork and her coauthors state: "[D]ifferential privacy by definition [protects against] re-identification."[70]

Similar to *k*-anonymity, differential privacy looks at a process of sanitization (e.g., when asked for your age, answer with your real age plus a random number from 1 to 10) and assesses how privacy preserving the output[71] is. The difference occurs, in part,[72] because differential privacy tells you how privacy preserving an output will always be, in a worst-case scenario, no matter what new privacy attacks are identified and no matter what new information an attacker learns. True enough, this protection comes at a cost—it is heavy handed, it does not apply to all scenarios, and it creates diminishing usefulness implications for the type of questions that may be answered[73]—but what it provides to the forest for the sake of the trees guarantees privacy like none other.

different things based on a similarity. In the database world, this often means two rows are merged together given similarity of an identifier, like a name. *See, e.g.*, Bellovin et al., *supra* note 57, at 4–5.

[69] A technique like *k*-anonymity was built to side-step joins, but that does not mean adversarial attacks using auxiliary data are completely off the table. A long line of literature following *k*-anonymity proved this. *See supra* notes 55–57 and accompanying text. In other words, that a technique is built specifically with one adversarial attack in mind, does not mean it has no weaknesses.

[70] Cynthia Dwork et al., *Exposed! A Survey of Attacks on Private Data* 61, 64–65 (2017) ("Reconstruction represents spectacular success on the part of the adversary, or, conversely, a spectacular failure of the putative privacy mechanisms. Tracing—that is, determining whether or not a specific individual is a member of a given dataset—is a much more modest adversarial goal."). In terms of tracing and reconstruction, reidentification would be somewhere in-between in terms of difficulty. *See infra* Section II.C.1.

[71] That standards like *k*-anonymity, *see supra* notes 58–60 and accompanying text, measure the privacy-preserving qualities of a *dataset*, rather than an *output* (as referenced above), does not displace the comparison here. Differential privacy is most easily thought of, didactically, in the query-response mode. *See infra* note 146 and accompanying text. A question is asked of a "recipe" which takes in data, processes it in some sanitizing way, and then sends it back out. That process produces "private" output, and although we are measuring the privacy preserving ability of the mechanism itself (rather than the output), we are nonetheless describing what can be learned, at most, by someone who wants to reidentify individuals in a sanitized output, regardless of whether that is a single answer (an output) or a dataset.

[72] *See infra* Part II for a full discussion.

[73] *See* Bambauer et al., *supra* note 16, at 723; *see also* Table 4 and accompanying text. Differential privacy is no panacea: it may not be right for all scenarios, it requires that certain types of questions be asked, and, by considering a worst-case scenario, it may be overly protective. *See, e.g.*, Matthew Fredrikson et al., *Privacy in Pharmacogenetics: An End-to-End Case Study of Personalized Warfarin Dosing*, *in* PROCEEDINGS OF THE 23RD USENIX SECURITY SYMPOSIUM 17 (2014), https://www.usenix.org/system/files/conference/usenixsecurity14/sec14-paper-fredrikson-privacy.pdf [https://perma.cc/CD4N-P8JV]. These limitations, however, may be a positive form of privacy friction. *See* Paul Ohm & Jonathan Frankle, *Desirable Inefficiency*, 70 FLA. L. REV. 777,

To more fully understand differential privacy, and how it may be "attacked" in a legal scenario, we must start with a few building blocks. The next Part provides the viewpoint from which differential privacy is most easily accessible, discussing the type of "noise"[74] differential privacy uses to measure privacy and explaining how the mathematical quantity of epsilon is used as an adjustable knob in the privacy–utility tradeoff.

II. BUILDING BLOCKS OF DIFFERENTIAL PRIVACY

Differential privacy requires a particular frame of mind. Similar to how, when considering the rule against perpetuities, it is important to take a step back and understand the perspective driving the rule,[75] differential privacy is most easily understood using the viewpoint from which it operates. This viewpoint may be grouped around three core concepts: (1) differential privacy focuses on "mechanisms" or "algorithms" (i.e., descriptions for how to accept some type of input, engaging with that input, and produce some type of output); (2) differential privacy is not a tool used to sanitize data, but is more like a standard, a statement about the privacy preserving abilities of a mechanism itself; and (3) differential privacy lives in a world of datasets, and produces its guarantees by measuring itself against a powerful adversary, quantifying how much information an attacker would, at most, be able to learn. The next section discusses each of these building blocks in turn.

A. Preliminary Cairns

For starters, differential privacy only concerns itself with *mechanisms*. A mechanism, broadly speaking, is a recipe, like a cooking recipe. Another term used for these recipes is a function or algorithm: a repeatable, consistent way of doing something that takes in a certain type of input and produces out a certain type of output.[76] It is best to think of this at the highest level possible: widget A

827–28 (2018) ("privacy is 'protected by the high cost of gathering or using' information, meaning that 'friction is . . . privacy's best friend.'" (citing William M Geveran, *The Law of Friction*, 2013 U. CHI. LEGAL F. 15, 15 (2013)).

[74] A terse term for this noise might be "lying"—a differentially private mechanism will provide back an "untruthful" answer to protect the "real" answer. *See generally* Bambauer et al., *supra* note 16. Nuance is lost, however, in that lying does not suggest a logical process behind the lie. Differential privacy is very particular about the type of noise it uses to maintain its guarantees, and calling this process "lying" is a disservice to that under-the-hood process.

[75] *See* Peter A. Appel, *The Embarrassing Rule Against Perpetuities*, 54 J. LEGAL EDU. 264, 264–66 (2004) ("Ask students what subject within property they hated most, and most will answer that it was the Rule Against Perpetuities. Indeed, it might rank as the most-hated doctrine studied in the first year of law school (although the *Erie* doctrine might give it a run for its money). Arcane in origin, difficult to understand and apply, unintuitive, and seemingly random in its effect, the rule brings together many of the difficulties that students have in adjusting to the rigors of legal study."). *But see* Shrutarshi Basu et al., *A Programming Language for Future Interests*, 24 YALE J.L. & TECH. 75, 79 (2022) ("Rather than using an existing programming language to write a program to model future interests, we treated the formalized, ritualized language of first-year Property conveyances as a programming language itself.").

[76] This is true even if the mechanism uses randomness. The function will take in some input, use randomness in a predictable, reputable way, and then produce out some output. Any mechanism that is called ε-differentially private works like this, from Google Chrome's RAPPOR to Facebook's

goes in and widget B comes out. For example, a mechanism for making a peanut butter and jelly sandwich would look like this:

```
Algorithm 1 Sandwich—Peanut Butter and Jelly
Input: Bread, Peanut Butter, Jelly

with (first slice of bread)                    ▷ pre-sliced bread
        add Peanut Butter to slice
with (second slice of bread)
        add Jelly to slice
Sandwich = merge first and second slice

Output: return Sandwich
```

Figure 1. An Algorithm for Making a Sandwich

We have an input of bread (two slices), peanut butter, and jelly. The algorithm takes this input, executes a sequence of operations (i.e., add peanut butter, add jelly, put the two together), and produces an output: a sandwich.

Differential privacy operates on mechanisms, like Algorithm 1 as shown in Figure 1. In fact, only mechanisms may be deemed ε-differentially private, not the results of the mechanism. We would not say the peanut butter and jelly sandwich (output) is ε-differentially private, but that the algorithm used to make the sandwich is ε-differentially private.[77] The ε part (Greek for epsilon) is discussed in Section II.D below. In short, it signifies how "private"[78] the output is. Also noteworthy is how differential privacy allows these functions to be made publicly available in a "don't-roll-your-own-crypto"[79] type of way. This allows for

"Full URLs" dataset. Solomon Messing et al., *Facebook Privacy-Protected Full URLs Data Set*, HARV. DATAVERSE (2022), https://doi.org/10.7910/DVN/TDOAPG.

[77] To be sure, the peanut-butter-and-jelly-sandwich algorithm is not, in fact, differentially private.

[78] This quick definition is purposefully ambiguous, given that it relies on the term *privacy*, which is famously ambiguous. Further clarification may be found in Section II.B *infra*.

[79] The idea is sometimes attributed to Bruce Schneier. *"Schneier's Law*,*"* SCHNEIER ON SECURITY: BLOG, https://www.schneier.com/blog/archives/2011/04/schneiers_law.html [https://perma.cc/3RVS-KDU6]. For an explanation, see AN INTRODUCTION TO CRYPTOGRAPHY 54 (1998) ("When I was in college in the early 70s, I devised what I believed was a brilliant encryption scheme. A simple pseudorandom number stream was added to the plaintext stream to create ciphertext. This would seemingly thwart any frequency analysis of the ciphertext, and would be uncrackable even to the most resourceful government intelligence agencies. I felt so smug about my achievement. Years later, I discovered this same scheme in several introductory cryptography texts and tutorial papers. How nice. Other cryptographers had thought of the same scheme. Unfortunately, the scheme was presented as a simple homework assignment on how to use elementary cryptanalytic techniques to trivially crack it. So much for my brilliant scheme. From this humbling experience I learned how easy it is to fall into a false sense of security when devising an encryption algorithm. Most people don't realize how fiendishly difficult it is to devise an encryption algorithm that can withstand a prolonged and determined attack by a resourceful opponent."); *see also* AUGUSTE KERCKHOFFS, LA

better mechanisms through open-source analysis and reproducibility. For instance, anyone may fully assess Google's RAPPOR system, which uses differential privacy in the Chrome web browser to check on intimate telemetry[80] details such as users' default homepages in their browsers.[81]

The second building block concerns differential privacy's role *not* as a tool of anonymization, but as a way of reasoning about mechanisms.[82] To be sure, differential privacy is not itself a tool for creating privacy. Unlike the methods of "generalization" (i.e., modifying data by generalizing it[83]) or "suppression" (i.e., modifying data by removing it[84]) differential privacy is a measure of privacy under a particular scenario. In many ways, it is like PII in that it ties to a general concept of privacy, but is not itself a way to achieve privacy.[85] For instance, COPPA defines "personal information"[86] as "individually identifiable information [(e.g., name, social security number, and e-mail address)] about an individual collected online"[87] while VPPA defines PII as any "information which identifies a person."[88] Neither one of these definitions provides a way to create privacy, yet, both provide a standard for evaluating privacy (or lack of privacy) in a specific setting. Similarly, the concept of differential privacy is not the narrow application of a tool to data; to understand the term in a working sense requires a setting, and this is where the third concept comes in.

Lastly, differential privacy lives in a world of datasets (i.e., tables of columnized information[89]) and adversaries (i.e., actors who wish to reidentify individuals in those datasets). In fact, the mathematical definition of differential

CRYPTOGRAPHIE MILITAIRE 8 (1883) (discussing how a cryptographic system should be secure regardless of whether its mechanics are known); Nissim et al., *supra* note 20, at 703–06.

[80] Telemetry data refers to general "use" information shared about a system. *See* Bolin Ding et al., *Collecting Telemetry Data Privately*, *in* 31ST CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS 3572, 3572 (2017) ("Counter data, e.g., daily app or system usage statistics reported in seconds, is a common form of telemetry.").

[81] *See* Erlingsson et al., *supra* note 15, at 1063.

[82] *See, e.g.*, Dan Feldman & Eldar Haber, *Measuring and Protecting Privacy in the Always-On Era*, 35 BERKELEY TECH. L.J. 197, 234 (2020) (describing differential privacy correctly as a standard).

[83] For example, the zip codes 93940, 93942, and 93943 may all be called "Monterey, California"—the data has been generalized, and therefore made more anonymous, by replacing all three datapoints with a single text statement.

[84] For example, the zip codes 93940, 93942, and 93943 may be replaced with 9394*, 9394*, and 9394*—the data has been suppressed, and therefore made more anonymous, by replacing the last digit of each zip code with an asterisk.

[85] *See generally* 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, 1819–36 (2011).

[86] Although *PII* is used as a term which applies to many data protective statutes, see *id.*, some statutes, like COPPA, may use phrases like "personal information" instead. *See* Children's Online Privacy Protection Act, 15 U.S.C. §§ 6501–6505 (Supp. IV 1998); 16 C.F.R. § 312.2 (2020) (defining "personal information" as "individually identifiable information about an individual collected online," including attributes like first and last name or social security number).

[87] 16 C.F.R. § 312.2 (2020).

[88] Video Privacy Protection Act of 1988, 18 U.S.C. § 2710(a)(3) (2006) (defining "personally identifiable information" as "information which identifies a person as having requested or obtained specific video materials or services from a video tape service provider").

[89] Literature on differential privacy often mixes the terms database and dataset. Here, we use the term dataset. *See supra* note 59 and accompanying text.

privacy makes no sense without these two elements. To better understand this world, consider the following table, which lists the names of individuals who ate certain types of cookies. This table would be deemed a "dataset."

Table 1. Original "Raw" Dataset

First	**Last**	**Birth-Year**	**Cookie Eaten**
Alice	Westminster	1984	Chocolate Chip
Bob	Kensington	2000	Gingersnap
Abigale	Westminster	1989	Chocolate Chip
Bob	Chelsea	2010	Gingersnap

The "adversary" here would be someone who tries to reidentify the individuals described by the dataset, which is a very simple task if Table 1 is shared in its raw form (i.e., just look at the table). Instead, what a data steward who owns the dataset and wants to "release and forget"[90] it into the wild may do is opt (even today[91]) for something like *k*-anonymity, which would create a sanitized-looking dataset.[92]

[90] *See* Ohm, *supra* note 11, at 1711–12.

[91] As the President's Council of Advisors on Science and Technology reported, "Anonymization of a data record might seem easy to implement. Unfortunately, it is increasingly easy to defeat anonymization by the very techniques that are being developed for many legitimate applications of big data." EXEC. OFF. OF THE PRESIDENT, REPORT TO THE PRESIDENT, BIG DATA AND PRIVACY: A TECHNOLOGICAL PERSPECTIVE 38–39 (2014), https://bigdatawg.nist.gov/pdf/pcast_big_data_and_privacy_-_may_2014.pdf [https://perma.cc/MLL5-F2HT]. Despite what we know about reidentification and the inadequacies of narrow definitions of PII, this understanding is still making the rounds today (e.g., a majority of the 50 states define PII narrowly when defining what actions do or do not trigger data breach protections). *See* Sara A. Needles, Comment, *The Data Game: Learning to Love the State-Based Approach to Data Breach Notification Law*, 88 N.C.L. REV. 267, 277–79 (2009) (finding that how a state defines PII hints at the state's overall objective for the purpose of a breach notification law).

[92] As Latanya Sweeney and her coauthors observed: "These findings suggest that there is something fundamentally flawed with ad hoc redactions of data. They fail to accurately account for the quality and nature of external information. Heavily redacted data may look anonymous, but it is not necessarily so." Latanya Sweeney et al., *Re-Identification Risks in HIPAA Safe Harbor Data: A Study of Data from One Environmental Health Study*, TECH. SCI., Aug. 28, 2017, at 1, 51–52, http://techscience.org/a/2017082801(click "Download" to access PDF) (arguing for increased sanitization in the HIPAA safe harbor redaction requirements).

Table 2. Sanitized-Looking Dataset

First	Last	Birth-Year	Cookie Eaten
-	Westminster	1980s	Chocolate Chip
Bob	-	-	Gingersnap
-	Westminster	1980s	Chocolate Chip
Bob	-	-	Gingersnap

For every row in the table, there are at least $k-1$ other rows in the table with the exact the same information (i.e., allowing Alice to "hide in the crowd" with Abigale by making the first and third rows in Table 1 identical). Practically, we are achieving a privacy "crowd" via k-anonymity by applying *suppression* (i.e., replacing a cell with "–") and *generalization* (i.e., making more general, like replacing the year 1984 with 1980s) all the while keeping our k goals in mind.

Although the above table looks sanitized, how confident should Alice be that no attacker could reidentify her? Should Alice be comfortable knowing that her data is deidentified? No.

A hypothetical adversary, Mallory, may have an easy time reidentifying Alice if Mallory happens to have access to another dataset (i.e., auxiliary information, Table 3) with full name and age information. All Mallory would have to do is search for all the Westminsters born in the 1980s to figure out that both Alice and Abigale ate chocolate chip cookies (i.e., PII may have been unveiled).[93]

[93] The assumption is based on the idea these two are the only two Westminsters born in the 1980s. Indeed, this is a toy example illustrating how an attacker may nonetheless learn from an anonymized dataset (i.e., it is not very sensible to leave a last-name column unmodified). A more typical, realistic attack on k-anonymity comes from the heterogeneity attack (i.e., learning what cookie Alice had eaten by knowing that Alice was born in the 1980s and has a last name of Westminster—the k-anonymized Table 2 does not hide this information well enough) or a background knowledge attack (i.e., assume the attacker knows only that the person they are attempting to learn about is deathly allergic to ginger, meaning that the only other option, in this dataset, is chocolate chip, and therefore the person must have eaten a chocolate chip cookie). *See infra* note 111 and accompanying text; *see also supra* note 56 and accompanying text (introducing these attacks).

Table 3. Original Plus Auxiliary Datasets

Original Dataset				*Auxiliary Dataset*		
First	**Last**	**Birth-Year**	**Cookie Eaten**	**Last**	**First**	**Birth-Year**
-	Westminster	1980s	Chocolate Chip	Alice	Westminster	1984
Bob	-	-	Gingersnap	Bob	Kensington	2000
-	Westminster	1980s	Chocolate Chip	Abigale	Westminster	1989
Bob	-	-	Gingersnap	Bob	Chelsea	2010

True enough, *k*-anonymity may preserve privacy if the *k* value is increased; the kernel in this example, however, is not how the adversary was able to uncover information, but that the strength of a privacy preserving standard is measured against an adversary who is assumed to always exist and always possess the goal of unveiling who is in the dataset. This is the perspective taken by differential privacy, motivating how it is technically defined.

In summary, differential privacy is a way of measuring the privacy of mechanisms acting on datasets in the face of an adversary. The following three Sections outline the core of why differential privacy works (randomness), what exactly differential privacy guarantees (the adversary), and how differential privacy uses randomness like a knob to increase or decrease privacy (epsilon). Together, these Sections represent the building blocks for the mathematical definition of differential privacy introduced in Part III. The next Section starts by introducing the randomness that differential privacy uses to purchase[94] privacy.

B. Why Differential Privacy Works: Randomness

Differential privacy works against an age-old quandary: How do you hide information while at the same time reveal information? For differential privacy, privacy is purchased[95] by avoiding real answers in a particular way, providing a veil of "plausible deniability" from the implications of a mechanism's output.

[94] True, differential privacy is not exchanging funds for privacy; however, the phrasing connotes the idea that privacy is not achieved for free. In the case of differential privacy, the cost is reduced utility owed to the use of randomness.

[95] *See supra* text accompanying note 94. Privacy is constrained by the concepts addressed at the beginning of Part I, *supra*. Indeed, with the possibly endless definitions of privacy going back to the fourteenth century, stating a singular definition of privacy and claiming that it is achieved may be a bit too far reaching.

Consider a hypothetical where Alice does not want anyone to know her real age. When asked her age, Alice responds with a random age near[96] her real age. The *recipe* or *mechanism* Alice uses has an input of "what is your age" and an output of {real age plus or minus a random number}.

To be sure, trusting Alice's response at face value, given that she uses the random-age mechanism, is unreliable; it is entirely possible and very likely that Alice has not provided her real age. True enough, knowing additional information about how Alice picks random numbers (to add or subtract from her real age) would help a detective (i.e., adversary) figure out exactly what Alice's real age is,[97] but assuming that random-age-choice information is off the table, Alice is free to proffer a responsive answer because her provided age is meaningless.[98]

In the same way, differential privacy relies on randomness to attain privacy. In fact, its inventors go so far as to state that "any *non-trivial* privacy guarantee that holds regardless of all present or even future sources of auxiliary information . . . . requires randomization."[99] On the other hand, using randomness, though effectuating privacy, degrades utility—what if we really did want to know Alice's real age?[100]

1. *Truth and Not-Truth: The Privacy-Utility Tradeoff*

If privacy cannot be attained without returning an unreal answer, then perfect privacy may be considered the opposite of perfect utility. We have privacy via randomness, but what if we also want utility?

Imagine trying to determine the ages of everyone in a particular neighborhood. If the Alice from our hypothetical, using the random-age generator, lived in this neighborhood, then the age-counts for this neighborhood would be inac-

[96] "Near" here is mathematically defined to be within a certain interval centering on Alice's real age.

[97] This is why differential privacy spends much of its time debating all possible results that could come out of the function. For instance, if the random-name generator only picked from three different names, including the name Alice, then we might say that Alice has less privacy when answering the question because an adversary knows that her name is only one of the three.

[98] For example, Alice's age information cannot be "joined" with other auxiliary information to uncover the real answer—the provided answer was *likely* fake to begin with. We say "likely" here because Alice is using a random number generator to come up with the number to add or subtract from her real age, and it may be possible that the random number generator includes 0 in its list of potential answers. Therefore, it is at least theoretically possible that Alice could respond with her real age even when using the range-age mechanism. *See also* Nathan Reitinger et al., *Is Cryptographic Deniability Sufficient? Non-Expert Perceptions of Deniability in Secure Messaging*, *in* PROCEEDINGS OF THE 44TH IEEE SYMPOSIUM ON SECURITY AND PRIVACY 274, 274–76 (2023) (discussing the implications of deniability in secure messaging systems).

[99] Dwork & Roth, *supra* note 1, at 226.

[100] Curtly, in the differential privacy setting, you would be out of luck; this question is not suitable for differential privacy. Differential privacy aims to protect individuals, and in this sense, if Alice's age mechanism were ε-differentially private with a "low enough" epsilon, then the random-age mechanism would *always* lie too much to allow you to figure out Alice's real age. This outcome is by design; differential privacy only allows "aggregate"-type queries, and this query is too specific. *See* Section II.D *infra* for a discussion of a "low enough" epsilon value.

curate, because Alice would most likely[101] lie about her real age. Alice's privacy is preserved, but the utility or accuracy of the overall count is harmed.

Elegantly, differential privacy uses the privacy-utility tradeoff to its advantage. By concerning itself with large-enough questions, differential privacy is able to play nice with Alice and the count, preserving Alice's desire to keep her true age private while also preserving the accuracy of the overall tally-count being close enough.

To see how this plays out, consider a neighborhood of 1,000 people and a specific question: How many people in this neighborhood are 33 years old? Assuming 100 of them are, in truth, 33 years old (including Alice), the real answer to this question is 100 (i.e., 10%). Given the lying mechanism[102] that Alice uses, however, the privacy-preserved answer would most likely[103] be 99 (i.e., 9.9%). Out of all the people who live in this neighborhood, 99 of them are 33 years old (including Alice's privacy preserved, unreliable answer).

The point here is not that 10% (i.e., the answer) is numerically close to 9.9%; rather, the point is that if the question concerns a large enough group, then the whole will be greater than its parts, the truth of the crowd outweighs Alice's lie. Imagine instead that the neighborhood only consisted of ten people. A lie here has an impressive impact on the outcome—adding an inaccurate 10% margin to any tally looking at age. This would be a large impact on utility.[104]

Differential privacy is powerful because it gets away with adding much more noise than simply one person out of the group lying—in fact, every person in the group receives the same insulation from the truth as Alice. For example, if we took an ASCII art picture[105] of a bike, modified all individual characters in the picture by flipping them blank (i.e., " ") or leaving them as is with 50% probability, then we would still be able to discern the overall picture, even though each character is insulated with a 50% chance of accurate–not accurate.

[101] If the random number pool Alice draws from when using her mechanism includes 0 then it is possible that Alice would respond with her real age, but, either way, if an attacker knew the mechanism was being used, then even the real age answer is unreliable—the attacker has to assume that the proffered age is inaccurate at least to some degree.

[102] As explained in *supra* note 74, Alice's lies are not the same as differential privacy's noise, though both return information which is untruthful. For Alice, a responsive answer is random, there is (necessarily) no way to predict which number is picked in the {real age ± random number} mechanism; for differential privacy, a responsive answer is not random, and (necessarily) adheres to a particular distribution. *See, e.g.*, *infra* Section III.A.

[103] *See supra* note 101 and accompanying text.

[104] If there is no one else in the neighborhood who is 33 years old, then with Alice's lie of "not 33" there are 0/10 or 0% of people who are 33. If Alice were to tell the truth, then this number magically becomes 10%—0% versus 10% is a large difference. If there was one other person in the neighborhood who was 33 years old, then either 10% (with lie) or 20% (without lie) of people are 33 years old. Again, a large difference given the size of this particular neighborhood.

[105] Technically, this picture of a bicycle was created using Unicode characters rather than pure ASCII characters. *See* David C. Zentgraf, *What Every Programmer Absolutely, Positively Needs to Know About Encodings and Character Sets to Work with Text*, KUNSTSTUBE (Apr. 27, 2015), https://kunststube.net/encoding/ [https://perma.cc/SA6P-WUX4].

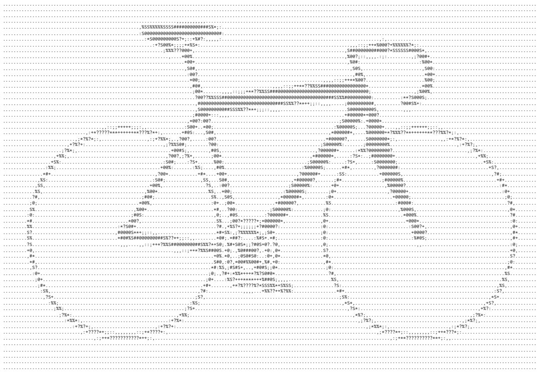

Figure 2. Original ASCII Image [106]

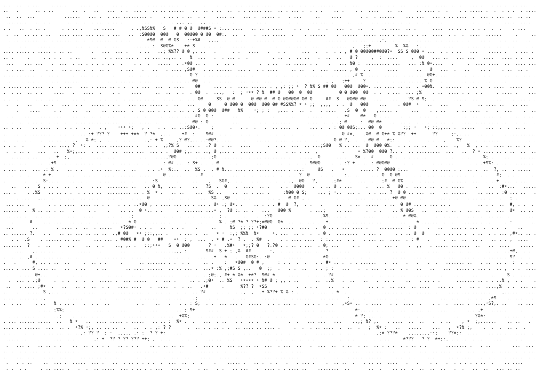

Figure 3. Sanitized ASCII Image[107]

So long as responsive answers (e.g., flipping each character or answering the "what is your age" question) are provided in a particular way, nothing will be learned about the individuals making up the group, but a fairly accurate some-

[106] This image comes from (public domain) from a black and white image of a bike. *See Vector Graphics of Bicycle*, FREE SVG (Mar. 24, 2014), https://freesvg.org/vector-graphics-of-bicycle7887 [https://perma.cc/CJ2V-Z63P]. The image is then processed into ASCII using a series of transformations on the raw image input. *See Convert Image or Text to the ASCII*-Art, ASCII-GENERATOR, https://ascii-generator.site [https://perma.cc/4EBU-A9PY]. Figure 2 shows the resulting ASCII image post-processing.

[107] Using Figure 2's image in text form, each character in the image was either modified with " " (i.e., a blank space) or left unchanged. The resulting picture nonetheless shows a bike. *See also* DIMACS/Northeast Big Data Hub Workshop on Privacy and Security for Big Data, *Utilizing Large-Scale Randomized Response at Google: RAPPOR and Its Lessons by Ananth Raghunathan*, YOUTUBE (Sept. 6, 2017), https://www.youtube.com/watch?v=tuOBz5AzivM [https://perma.cc/Y9CG-P9XR] (discussing Google's RAPPOR system by first teaching differential using a similar example on a Mona Lisa ASCII art picture); *see also* Priyanka Mathikshara, *(Local) Differential Privacy for Dummies :D*, MATHIKSHARA.COM (Aug. 4, 2020), https://www.mathikshara.com/post/local-differential-privacy-for-dummies [https://perma.cc/Q956-XN9B].

thing will be learned about the group as a whole. Stated in more mathematical terms, any output of a differentially private mechanism is nearly as likely regardless of whether one individual was "in" the dataset or not.[108]

C. Adversarial Perspective

Taking a step back, it is important to note the why behind differential privacy's use of noise to provide inaccurate answers. The why here comes directly from the historic perspective of reidentification attacks: deidentification talks more than it walks.[109]

Differential privacy takes a nod from the failings of Netflix Prize and the AOL search log[110] by leaving room for the possibility that someone may try and use any and all auxiliary information (i.e., information unbounded by the instant dataset) in a hodgepodge aimed at reidentification. And this goes beyond the practical attack Professor Sweeney persuasively demonstrated in 1997 (e.g., taking public voter list records and joining[111] them with deidentified[112] medical records). Instead, differential privacy directly addresses the means used to effectuate those attacks: reconstruction attacks.[113]

Simply speaking, reconstruction attacks take advantage of the fact that computer time and human time are different. One of the most magical parts of

[108] *See generally* Graham Cormode et al., *Privacy at Scale: Local Differential Privacy in Practice*, *in* SIGMOD '18: PROCEEDINGS OF THE 2018 INTERNATIONAL CONFERENCE ON MANAGEMENT OF DATA 1655 (2018), https://dl.acm.org/doi/pdf/10.1145/3183713.3197390 [https://perma.cc/5N54-U9UH]; *see also* Graham Cormode et al., *Privacy at Scale, Part B*, Lecture Slides, https://sites.google.com/view/kdd2018-tutorial/home/slides (Part B) [https://perma.cc/22S4-86RQ].

[109] In other words, a dataset may look well sanitized in practice, but throwing powerful computers at it seem to melt away those protections entirely. *See* Brief for U.S. Dep't of Commerce as Amici Curiae Supporting Respondents, Alabama v. U.S. Dep't of Commerce (No. 3:21-CV-211-RAH), at *12, https://www.brennancenter.org/sites/default/files/2021-04/Amicus%20Brief_data privacyexperts_%202021-04-23.pdf [https://perma.cc/2TVK-6QAJ] ("At this point, re-identification of 'anonymized' data is taken for granted by the academic privacy community. It is no longer an open research question.").

[110] Several widely publicized reidentification examples come from the mid-2000s, including both the Netflix Prize affair and the AOL search log debacle. Both examples stem from releases of data assumedly anonymized, but later found to be re-identifiable. *See* Boris Lubarsky, *Re-Identification of "Anonymized" Data*, 1 GEO. L. TECH. REV. 202, 208–12 (2017) (discussing these examples at a high level). *See generally* Ohm, *supra* note 11, at 1716–22.

[111] *See* RAMEZ ELMASRI & SHAMKANT B. NAVATHE, FUNDAMENTALS OF DATABASE SYSTEMS 251 (7th ed. 2016) ("The JOIN operation, denoted by ⋈, is used to combine related tuples from two relations into single "longer" tuples. This operation is very important for any relational database with more than a single relation because it allows us to process relationships among relations."); *see also* RAGHU RAMAKRISHNAN & JOHANNES GEHRKE, DATABASE MANAGEMENT SYSTEMS 107 (3d ed. 2003) ("The *join* operation is one of the most useful operations in relational algebra and the most commonly used way to combine information from two or more relations. Although a join can be defined as a cross-product followed by selections and projections, joins arise much more frequently in practice than plain cross-products.").

[112] *See* Latanya Sweeney, *Simple Demographics Often Identify People Uniquely* 1 (Carnegie Mellon Univ., Data Priv. Working Paper No. 3, 2000), https://dataprivacylab.org/projects/identifiability/paper1.pdf [https://perma.cc/TGF8-XNTL]).

[113] *See* Arvind Narayanan & Vitaly Shmatikov, *Robust De-Anonymization of Large Sparse Datasets*, *in* PROCEEDINGS OF THE 2008 IEEE SYMPOSIUM ON SECURITY AND PRIVACY 111, 112 (2008), https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4531148.

computing is a computer's ability to execute computations with blazing speed (e.g., *variable a* = 1 + 1) and remember those computations (e.g., *variable b* = 1 + *a*) in a useful way (e.g., *variable b* is 3).[114] What this means for a chess master like Garry Kasparov is bad news, at least when it comes to winning,[115] because as long as a problem can be represented mathematically, then a computer can blindly work on it for what would be considered decades of human time—that is, Kasparov lost because Deep Blue checked many, but not all, possible combinations of chess moves that could be made[116]—but are merely seconds in computer time[117] (also known as cheating).

The same time difference is leveraged in reconstruction attacks by knowing an output (i.e., you are given an answer, like, the mean age of my classmates is 24) and finding all possible combinations of numbers that could lead to that output (e.g., (23 + 25) ∕ 2).[118] This may seem impossible given a complicated output, but with unlimited guessing and nearly unlimited storage capacity, it accords with logic to say that the answer will eventually come to the fore.

Two key presuppositions can be learned from the reconstruction attack. First, some combinations of numbers are more likely than others. For example, it is unlikely that, if the average age of a group of classmates is 24, and if I know

[114] In short, this is because computers are universal machines. *See, e.g.*, Ian Watson, *How Alan Turing Invented the Computer Age*, SCI. AM. (Apr. 26, 2012), https://blogs.scientificamerican.com/guest-blog/how-alan-turing-invented-the-computer-age/ [https://perma.cc/8QSQ-V8E7].

[115] There is a long history of fairness in competition, typically requiring a win to be won fairly, something that may be called into question given the unfair abilities of computing. *See, e.g.*, Avila v. Citrus Community College Dist., 41 Cal. Rptr. 3d 299, 302 (2006) (discussing the legality of a "beanball": in baseball, a pitch that targets a batting player's head).

[116] For an updated version of Deep Blue, see alphaGo David Silver et al., *Mastering the Game of Go with Deep Neural Networks and Tree Search*, 529 NATURE 484, 488 (2016) ("In this work we have developed a Go program, based on a combination of deep neural networks and tree search, that plays at the level of the strongest human players, thereby achieving one of artificial intelligence's 'grand challenges'); ALPHAGO (Netflix 2017). Alphabet Inc. followed alphaGo up with a self-learning version, a difference in kind from Deep Blue and alphaGo: alphaGo Zero). *See* Dawn Chan, *The AI That Has Nothing to Learn from Humans*, ATLANTIC (Oct. 20, 2017), https://www.theatlantic.com/technology/archive/2017/10/alphago-zero-the-ai-that-taught-itself-go/543450/ [https://perma.cc/V3CG-RSTA] ("You have to be ready to deny a lot of the things that we've believed and that have worked for us.").

[117] CPUs crunch numbers very quickly, without getting tired. Imagine if you had to calculate many (not all, the space is quite large) chess moves' outcomes on every single chess move. Sure, the problem gets easier as the game goes on, but most of the game is built on millions of possible moves. Gil Press, *The Brute Force of IBM Deep Blue and Google DeepMind*, FORBES (Feb. 7, 2019, 9:18AM) https://www.forbes.com/sites/gilpress/2018/02/07/the-brute-force-of-deep-blue-and-deep-learning/?sh=6dafb3be49e3 [https://perma.cc/9HE3-2XC2] ("Deep Blue was an example of so-called "artificial intelligence" achieved through "brute force," the super-human calculating speed that has been the hallmark of digital computers since they were invented in the 1940s. Deep Blue was a specialized, purpose-built computer, the fastest to face a chess world champion, capable of examining 200 million moves per second, or 50 billion positions, in the three minutes allocated for a single move in a chess game.").

[118] Cynthia Dwork, *Ask a Better Question, Get a Better Answer: A New Approach to Private Data Analysis*, *in* DATABASE THEORY—ICDT 2007: 11TH INTERNATIONAL CONFERENCE 18, 18–20 (Thomas Schwentick & Dan Suciu eds., 2006); Irit Dinur & Kobi Nissim, *Revealing Information While Preserving Privacy*, *in* PODS'03: PROCEEDINGS OF THE TWENTY-SECOND ACM SIGMOD-SIGACT-SIGART SYMPOSIUM ON PRINCIPLES OF DATABASE SYSTEMS 202 (2003), https://dl.acm.org/doi/pdf/10.1145/773153.773173.

that there are nine classmates in the class, then the ages of the nine classmates are 200, 9, 1, 1, 1, 1, 1, 1, 1—though this makeup does produce an average age of 24. Second, hunches about real answers will improve over time if repeat questions are permitted.[119]

In the first case (i.e., likely combinations), the questions you may ask a function are not all created equal, particularly with respect to an output. Some questions have specific answers, others have general answers. What this means for a reconstruction attack is that some answers are more reconstructable than others because only a few combinations produce the particular output. Differential privacy takes this into consideration when deciding how much noise to add to a function's output. In fact, differential privacy, using the common Laplace method, considers explicitly the maximum range of values there might be when assigning noise. As Part III more fully explains below, the fewer combinations there are, the more noise is needed.

And in the second (i.e., repeat questions), if we are playing the guess-this-input-given-that-output game, from the perspective that each time we see an output we come up with a list of combinations that produce that output, then it is easy to see how repeating questions allows for a paring down of possible combinations. In a brute force type of way, if we ask the same question over and over again, we will eventually find the real answer, regardless of the inaccuracies reported over time.[120] For example, if you give me a random answer which deviates slightly from the real answer each time I ask for it, all I need to do is average the random answers to get better and better hunches of the real answer. If I ask Alice "what is your age" over and over again, and Alice says: 33, 34, 30, 31, 33, 33, 37 then I might start to get the suspicion that Alice's real age is 33.

In a more nuanced sense, each time we reconstruct the possible inputs to produce an observed output, we are producing a set of combinations, and we know that the space between all of these combinations is where the real answer lies. In this way, the space gets smaller and smaller as we ask more and more questions. This is why you may have heard rumors of a privacy budget.[121] The

[119] This is known as the database reconstruction theorem. *See* Dinur & Nissim, *supra* note 118 (finding that accurately revealing *any* information, over time, no matter how small, from a dataset will eventual destroy all "privacy" the dataset had—i.e., only publishing statistical summaries will betray privacy, eventually); *see also* Dwork et al., *supra* note 70, at 64 ("There is by now a rich literature showing that any mechanism providing overly accurate answers to too many linear queries is blatantly nonprivate, meaning that it succumbs to a reconstruction attack. Indeed, there is a single attack strategy that succeeds against all such overly accurate answering of too many queries. Here, 'too many' is quite small (e.g., only *n* queries) and 'overly accurate' means having fractional error on the order of $o(1/\sqrt{n})$.")

[120] The "it depends" here turns on whether the mechanism generates answers independently (i.e., the case explained above) or remembers questions asked and returns the previous answer provided (i.e., a more common case in actual differential privacy deployments).

[121] *See, e.g.*, Liu, *supra* note 18, at 497; Myres & Nelson, *supra* note 16, at 137; Andrea Scripa Els, *Artificial Intelligence as a Digital Privacy Protector*, 31 HARV. J.L. & TECH. 217, 220 (2017) ("How much of an impact the data must have on the query to be excluded—and by extension how likely it is that a query would lead to personal identification—depends on a 'privacy budget' set by the holder of the data, which defines how much information leakage is considered acceptable.").

budget runs out the more you ask questions. That said, this is a well-known aspect of differential privacy, and one that can be controlled.

Differential privacy overcomes both how much noise and repeat questions with a adjustable knob known as epsilon. Knowing how to turn this knob depends, essentially, on how privacy sensitive an output is. The following section discusses how epsilon responds to these two issues in more depth.

D. What the Knob Means—Epsilon

Epsilon is the most important part of differential privacy.[122] The reason for this, however, may not be what you are thinking.

1. *Non-Contextual Epsilon*

A naïve way to think about epsilon would be to consider it the amount of noise that is added within a mechanism. A lot of noise is added with a small epsilon value (e.g., .01) and almost no noise is added with a large epsilon value (e.g., 10). If the output is privacy sensitive, like the answer to a sensitive question such as "have you ever had an abortion," then you will likely want more buffer room between the real answer and the mechanism's output; but if the question is not considered very sensitive, such as "do you like pizza," then you might use a larger epsilon value, meaning the provided answer is more likely close to the truthful answer.

This is how epsilon works in a mathematical sense, with more nonsensical output associated with low epsilon values and basically real outputs associated with high epsilon values, but the problem with this understanding is that it has no context. What does an almost-real output mean? Why should anyone be comfortable with a mechanism that used a small epsilon value but nonetheless outputs a number close to the real answer? Context is necessary and context for differential privacy comes from the adversarial perspective.

2. *Contextual Epsilon: Bounding*

Epsilon matters is because it bounds the threat of privacy loss. Epsilon says: this output (i.e., mechanism's answer) is no more meaningful than an increase in some percentage of a belief that it is correct.[123] In other words, confidence in

[122] Interestingly, it is also relatively scantly addressed in the technical literature, at least when it comes to setting an epsilon value. This is likely because setting epsilon requires considering the particular setting a mechanism is being applied to, which cannot be done a priori. *See* Cynthia Dwork et al., *Differential Privacy in Practice: Expose Your Epsilons!*, 9 J. PRIV. & CONFIDENTIALITY, no. 2, 2019, at 1 (arguing for an epsilon repository); *see also* Mary Anne Smart et al., *Understanding Risks of Privacy Theater with Differential Privacy*, 6 PROC. ACM HUM. COMPUT. INTERACTION, no. CSCW2, 2022, art. no. 342, at 1, 1 (2022), https://dl.acm.org/doi/pdf/10.1145/3555762 ("In implementations of differential privacy, certain algorithm parameters control the tradeoff between privacy protection for individuals and utility for the data collector; thus, data collectors who do not provide transparency into these parameters may obscure the limited protection offered by their implementation. . . . ").

[123] True enough, this is not how statutes regulate privacy risk, neither is it how typical users think of privacy risk, and this is why a proxy value is needed to translate what differential privacy provides—bounded loss—to what data protective statutes require—a certain threshold of appropriate reidentification risk. *See infra* Section III.B.2; *see also* Rachel Cummings et al., *"I Need a Better*

a guess at the real answer, when seeing the output of a mechanism, will never go beyond the limit set by epsilon. Your answer of "yes I have had an abortion" may only be 2% more likely to be the true answer, which is likely not high enough for me to trust that it is the true answer. Epsilon says that someone seeing your answer to this question will never have more than a 2% confidence boost that this is the real answer.[124]

In more concrete terms, differential privacy guarantees that an attacker, with some predefined, best-guess idea at an outcome, who views the results of a mechanism, cannot learn more, in percentages, than is controlled by epsilon.[125] For low values of epsilon, this means that the attacker's initial suspicion (e.g., 50%) will not change very much, probability wise, after seeing the mechanism's output (e.g., from 50% to 52%). For high values of epsilon, this means the attacker's initial belief that an outcome is real (e.g., 50%) may grow substantially after seeing an output (e.g., from 50% to 75%). The same is true regardless of the level of initial suspicion. If an attacker knew the real answer was a number between one and ten, then attacker has a 10% guess out of the gate—but if epsilon was set to be high, then the attacker may, after seeing an output, believe there is a 95% chance that the observed output is real (i.e., believe that this specific value in a range of possibilities is likely to be the real answer with a 95% chance).[126] And this is why differential privacy is only meaningful in terms of the particular epsilon a mechanism wields—a high epsilon means that there is practically no privacy, the results of the function are almost-but-not-quite right; a low epsilon means that there is practically no usefulness to the data, the results of the function are too incorrect to be useful. This is why we do not call a mechanism (i.e., recipe) differentially private, but ε-differentially private. The following Part takes this understanding one step further by unveiling the mathematical definition of differential privacy.

III. DEFINITION AND STEP ONE

A more formal[127] definition of differential privacy looks like this:

Description": An Investigation into User Expectations for Differential Privacy, *in* CCS '21: PROCEEDINGS OF THE 2021 ACM SIGSAC CONFERENCE ON COMPUTER AND COMMUNICATIONS SECURITY 3037 (2021), https://dl.acm.org/doi/pdf/10.1145/3460120.3485252 (investigating user opinions on differential privacy across a variety of metrics).

[124] *See infra* Figure 6 (showing how various values of epsilon impact an attacker confidence that a provided answer is the true answer). An answer to this question (have you ever had an abortion) is either yes or no. Depending on the mechanism under consideration, epsilon may tell something like this: the attacker may gain at most a 2% boost in confidence in an answer—there is a 52% chance that the observed answer is the real answer.

[125] Truthfully, it is e, Euler's number, with an exponent of epsilon—e^{ε}—but this simplification is permissible for now. Additionally, differential privacy compares this value against a multiplicative measure of similarity between two similar datasets (see the definition in Part III *infra*). For more detail, see Salil Vadhan, *The Complexity of Differential Privacy*, *in* TUTORIALS ON THE FOUNDATIONS OF CRYPTOGRAPHY 347, 354–58 (Yehuda Lindell ed., 2017).

[126] Perhaps it is very, very unlikely that a "real" answer is 1, but the output is 1, which means the attacker has "learned" a lot, in terms of confidence, that the observed output of "1" is the real answer.

[127] This version omits certain details which are needed to mathematically prove the equation, but are unnecessary for a basic understanding.

$$\frac{\mathbb{P}[Mechanism(input_{dataset1}) = output]}{\mathbb{P}[Mechanism(input_{dataset2}) = output]} \leq e^{\varepsilon}$$

Although this equation may appear jarring, Part I covered its most difficult parts. For notation, the $\mathbb{P}$ in both the numerator (top) and denominator (bottom) simply mean "the probability"; in this case, the probability that version one or version two of the mechanism's input will have a particular (same) output, which has to be less than or equal to *e*, a number,[128] raised to ε, the epsilon value discussed in Section II.D above. If the numerator were one and the denominator were two, then the equation would simply look like this: $1/2 \leq e$^ε. The number *e*, Euler's number, may be simply thought of as approximately 2.71828 (i.e., $1/2 \leq 2.7$^ε).

The two datasets[129] of the mechanism's input (*dataset1* and *dataset2*, numerator and denominator, respectively) are meant to capture the situation where the data the function operates on differs in a small way, while using the same mechanism. For example, using Table 1's "cookie eaten" column (mechanism: name the type of cookie eaten), *dataset1* would be the a dataset with someone eating a gingersnap and *dataset2* would be the same dataset, but this time without that person eating the gingersnap.[130] Differential privacy looks at the problem this way because it attempts to capture the reconstruction attack: If my best guess combination to produce an output similar to the mechanism's output is as good as I can get—i.e., my combination which produces this output is only one missing piece away—then what does that mean for privacy loss?

In summary, at a high level, the mathematical definition of differential privacy requires that a mechanism's output (e.g., cookie count) on a dataset (e.g., one gingersnap eaten by the individual) be close to the mechanism's output (e.g., cookie count) on a similar dataset (e.g., zero gingersnaps eaten by the individual). Why differential privacy is able to offer a "privacy guarantee" is because it is able to define close mathematically: the left side of the equation (i.e., fraction) must be equal to or smaller than the right side (i.e., Euler's number raised to epsilon). In other words, the mechanism makes a similar statement both

[128] This number is known as Euler's number and shows up across a variety of disciplines. *See* Stefanie Reichert, *e is Everywhere*, 15 NATURE PHYSICS 982, 982 (2019) (discussing Euler's number). For our purposes, it may be helpful to think of the number as simply ~2.7.

[129] Truthfully, differential privacy requires that this equation holds for all datasets. However, it may be easier to imagine a case where the two datasets are similar, *but for* one particular datum, a case which is most likely to produce a fraction equal to or larger than the value of epsilon.

[130] More specifically, the person eating the gingersnap would be present (i.e., "in") the first dataset, and not present (i.e., not "in") the second dataset. In this way, the two datasets differ in regard to a single record. JOSEPH P. NEAR & CHIKÉ ABUAH, PROGRAMMING DIFFERENTIAL PRIVACY 19 (2022) ("Two datasets are considered neighbors if they differ in the data of a single individual. . . . The important implication of this definition is that [the mechanism's] output will be pretty much the same, with or without the data of any specific individual. In other words, the randomness built into [the mechanism] should be 'enough' so that an observed output from [the mechanism] will not reveal which of [*dataset1*] or [*dataset2*] was the input. Imagine that my data is present in [*dataset1*] but not in [*dataset2*]. If an adversary can't determine which of [*dataset1*] or [*dataset2*] was the input to [the mechanism], then the adversary can't tell whether or not my data was present in the input—let alone the contents of that data.").

with and without the data from the person who ate a gingersnap—the gingersnap lover's data must be, in some ways, meaningless.

For a more technical explanation, which is helpful didactically, we can take a look at a mechanism that has been around for a long time: randomized response.[131] Indeed, any mechanism, including those created before the invention of differential privacy, may be analyzed with the lens of differential privacy. Differential privacy did not invent privacy preserving algorithms, it is simply a means of measuring one type[132] of privacy loss that an algorithm encumbers. If a mechanism has some measure of randomness, then the mechanism may[133] be proved to have a calculable ε, representing an ε-differentially private algorithm. Randomized response, in the setup given below, has an ε value of ~1.098.

A. Mechanism—Randomized Response: A Teaching Tool

Imagine we are using the following algorithm:

[131] To be sure, our understanding of differential privacy so far, and the eventual two-step test discussed in Part IV *supra*, applies to situations beyond randomized response; randomized response simply presents a simplistic way to understand differential privacy. The same is true for our two-step test's direct applicability to differential privacy in the "query" mode—that is, a question is asked of a mechanism and a privatized response is provided in return. There are many ways to achieve differential privacy (e.g., local differential privacy) and we picked what we thought was the easiest from a didactic standpoint.

[132] To be sure, differential privacy is one way, a mathematical way, of looking at privacy. This is not to say no other means exist to describe privacy or that differential privacy fully captures an understanding of privacy. *See* Dwork et al., *supra* note 70, at 77 ("The limits imposed by reconstruction and tracing attacks are absolute: no mechanism protecting against reconstruction and tracing can introduce less noise than is required to stymie the attacks discussed earlier. However, there are other adversarial goals, such as learning the sickle cell status of a specific individual, that do not require reconstruction, re-identification, or tracing, and each of these new goals may have its own set of attack strategies. A privacy solution that rules out reconstruction and tracing may not rule out attacks satisfying these other goals. The cryptographic approach to this dilemma is to first define privacy and then provide techniques that provably satisfy this definition. If the definition is too weak, in that it fails to protect against an important class of adversarial goals, it can be strengthened and new algorithms designed. The advantage to the definitional approach is that, because the definitions are getting stronger, progress is made. Differential privacy was first proposed in 2006 and so far has not required strengthening.")

[133] This is not yet a proved assertion! The idea, however, has grounding in the literature. *See* Dwork & Roth, *supra* note 1, at 216.

```
Algorithm 2 Randomized Response
Input: Are you a member of the Communist Party?

flip a coin                          ▷ 50% chance heads or tails
    if tails:
        Response = tell the truth
    if heads:
        flip a second coin
        if heads:
            Response = "Yes"
        if tails:
            Response = "No"

Output: return Response
```

Figure 4. Algorithm 2

This mechanism[134] has an input of a question and an output of the answer to that question.[135] The mechanism uses coin flips to insulate a respondent's secrets, similar to Alice's random-age generator.[136] If the coin lands tails, then the question is answered truthfully, but if the coin lands heads, then the question is answered true or false depending on another coin flip. In this way, the mechanism buys privacy with the fifty-fifty–tails-heads odds.[137]

We may analyze[138] this mechanism by noting all possible outcomes. We can then find the best-case scenario for an attacker to learn as much as possible

[134] *See generally* Dwork & Roth, *supra* note 1; Daniel Kifer, *Introduction to Differential Privacy*, NBER (July 17, 2020), https://www.nber.org/sites/default/files/2021-01/KiferIntroduction.pdf [https://perma.cc/72TN-NYZY]; Mark Bun, *A Teaser for Differential Privacy*, CS.PRINCETON (Dec. 8, 2017), https://www.cs.princeton.edu/~smattw/Teaching/521fa17lec22.pdf [https://perma.cc/5UZU-UZVY]; Damien Desfontaines, *Differential Privacy in (A Bit) More Detail*, DESFONTAIN.ES, https://desfontain.es/privacy/differential-privacy-in-more-detail.html [https://perma.cc/C6TX-2J5K] (Feb. 20, 2019).

[135] "Sensitive" here being illegal behavior. Randomized response has been used to assess the state of abortions pre-*Roe v. Wade* and tax evasion. *See* James R. Abernathy et al., *Estimates of Induced Abortion in Urban North Carolina*, 7 DEMOGRAPHY 19 (1970); Jodie Houston & Alfred Tran, *A Survey of Tax Evasion Using the Randomized Response Technique*, 13 ADVANCES TAXATION 69 (2001); Stanley L. Warner, *Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias*, 60 J. AM. STAT. ASS'N 63 (1965); Bernard G. Greenberg et al., *The Unrelated Question Randomized Response Model: Theoretical Framework*, 64 J. AM. STAT. ASS'N 520 (1969).

[136] *See supra* Section II.B.

[137] *See* Dwork & Roth, *supra* note 1, at 255.

[138] This is the Bayesian method of analysis. That is a way of thinking about differential privacy. *See* Desfontaines, *supra* note 134.

from the response.[139] Notably, an attacker's best-case scenario is the one which has the most likely outcome.[140]

As Figure 4 shows, only two possible outcomes for Algorithm 2 exist: yes or no. You either are or are not a member of the Communist Party. Given the definition of differential privacy from above, we may consider the case where $input_{dataset1}$ is a yes—a "dataset" with a person who would answer yes ("real answer") to the question being asked. Therefore, the only other possibility for $input_{dataset2}$ is a "no," and the person would answer "no" as the real answer. Stated otherwise, what is the probability (numerator) of a "yes" (output) with someone whose real answer is yes, and what is the probability (denominator) of a "yes" (output) with someone whose real answer is no—we are trying to figure out all the ways a yes occurs, letting us know what the probability of seeing a yes is. This gives us the probability of a yes in the best case for the attacker (i.e., the most we can learn when we see the output of the randomized response algorithm—the worst case for privacy).

$$\frac{probability\ of\ a\ yes\ given\ a\ real\ answer\ of\ yes}{probability\ of\ a\ yes\ given\ a\ real\ answer\ of\ no} \leq e^{\varepsilon}$$

For the numerator (i.e., top line), a yes can occur with a 50% chance of being truthful (line 2, Algorithm 2) *or* a 50% chance of landing heads and then a 50% chance of landing on heads again (line 6, Algorithm 2). Together, this is a 75% chance (.50 + (.50 * .50)). For the denominator (i.e., bottom line) to be yes with a real no answer, the first flip must be heads (line 4, Algorithm 2) and the second flip must also be heads (line 6, Algorithm 2). This case happens with a 25% chance (.50 * .50). Therefore, assuming we are talking about the probability of a yes in general, we can say that there is a (.75 / .25) fraction that this occurs, or a whole number of three. We plug this into the differential privacy equation:

$$3 \leq e^{\varepsilon}$$

[139] In this case, we are revealing the mechanism's epsilon value instead of setting it, given that the algorithm is stated as fact. Typically, given the popular Laplace method used in differential privacy, epsilon is an input variable to the question. In this way, epsilon may be modified to meet a particular privacy-utility balance. The randomized response protocol described in Figure 4's Algorithm 2, however, uses set coins to achieve randomness. Therefore, Algorithm 2's epsilon value is more accurately "identified" via proof rather than modified on an ad hoc basis.

[140] By analogy, if a ball was hidden underneath one of two buckets, bucket #1 with a 20% chance and bucket #2 with an 80% chance, and you had to guess which bucket the ball was in, you would logically pick the second bucket. Time-flipping this, if the ball is out and you had to guess which bucket it came from, the best-guess is bucket #2. Likewise, given that we are analyzing all possible inputs given a particular output, the output with the highest probability will be the "best case" for the attacker; if this outcome is returned by the mechanism, then there is a higher chance that is the "real" answer.

A math trick allows us to rephrase this statement to make it cleaner: the natural logarithm of three must be less than or equal to epsilon. This number, rounded, is approximately 1.098.

This results in the worst-case scenario for the respondent (i.e., highest probability of seeing a yes), meaning that this value sets our epsilon in this particular algorithm. Algorithm 2 is therefore deemed (1.098)-differentially private. Importantly, as Section II.D.2. emphasized above, this is an expression regarding the bounds of what an attacker may learn when seeing the output of a mechanism.

B. Differential Privacy Takeaways

Taking a step back and focusing on the task at hand—translating differential privacy into something legally meaningful—a problem is found with the previous Section's closing statement: it is legally meaningless.[141] Data regulation does not speak directly to differential privacy and the idea of bounded privacy loss; instead, statutes regulating data require data to be kept confidential (i.e., not shared) if the data is considered PII.[142] And although one way to transform PII into non-PII is to sanitize it, it is difficult to know exactly how sanitized resulting outputs are and how much sanitization a statute requires. Therefore, what needs to be found is a common measurement between what a statute deems sufficient sanitization and what a mechanism technically provides.

Luckily, differential privacy offers one of the most applicable, system-to-system comparisons for privacy that exists: epsilon—privacy by any other name.[143] If properly framed, the attributes inherent to differential privacy allow it to be consistently and repeatably applied to legal questions. In this way, what a mechanism technically provides may be rephrased, legally speaking, as reidentification risk. Before diving into possible options for framing differential privacy in terms of a reidentification risk, however, we must first address the fact that, mathematically speaking, differential privacy says nothing about reidentification risk.[144]

1. *Reidentification: Appropriate Overprotection*[145]

To clarify, reidentification occurs when an individual's data found within a dataset is no longer anonymous. An attacker is able to point at a record and say

[141] Previous work has attempted to make this translation easier, but a generally applicable lens may nonetheless be beneficial. *See, e.g.*, Nissim et al., *supra* note 20 and accompanying text.

[142] *See infra* notes 173–178 and accompanying text.

[143] *See, e.g.*, Dwork et al., *supra* note 122, at 4.

[144] True enough, differential privacy does protect against reidentification risk, but, speaking technically, the equation in Part III's introduction does not include a statement about the reidentification of individuals. *See id.* at 2–3 (discussing reidentification risk).

[145] We use the term "reidentification risk"—and not something like "disclosure risk"—because this is the term which is more likely common to a legal audience; in fact, this is a term which has already been picked up by several courts debating issues of data sanitization requirements. *See infra* Section IV.A. We note, however, that the term is, in some ways, lacking. For example, assuming a nearly nonexistent privacy budget (i.e., epsilon) and a query of "how many people have Crohn's disease" would not, on its face, reidentify someone. Learning the truth of this question is not, per se, a reidentification. That said, it does increase the *risk* of reidentification, which is what

"this is your data," or, for differential privacy in the query setting discussed so far,[146] see the output of a mechanism (e.g., did Abigale eat a chocolate chip cookie—yes) and know it is real.[147] This is a spectacular failure for a dataset—game over for an individual.

Importantly, differential privacy does protect against reidentification attacks, but it also protects against other types of attacks as well. Differential privacy must protect against all types of attacks for its guarantee to hold.[148] For example, in a successful tracing attack, which is covered by differential privacy's protection guarantee, an attacker merely learns whether an individual is in a dataset, not what the individual's data is (e.g., is Abigale in the "cookies eaten" dataset).[149]

The problem is that summarizing differential privacy in terms of reidentification risk inherits this overprotection, and what this means for our legal comparator, introduced in Section III.C below, is that we will be necessarily overprotecting data. That said, hinging protection on an overinclusive definition has several advantages.

First, using overprotection provides breathing room to an otherwise uneasy ask—releasing protected data into the wild. Understanding that the measure of reidentification risk borne from a differential privacy mechanism assumes a worst-case scenario gives balance to that proposition. Second, and more important, this amount of overprotection is necessary to prevent new, currently unknown attacks from degrading current standards of sanitization (i.e., differential privacy is futureproof). As discussed in Section IV.A.2 below, a thorn for

statutes aim to regulate. Our two-step test using the "guess difference" value aims to provide a roadmap for interpretation—a translation—of differentially private mechanisms in a legal setting, not create a new, rigorous, mathematical way of measuring the risk that mechanisms encumber (i.e., something like the disclosure risk of a mechanism). *See supra* Section III.B.

[146] Differential privacy has many different implementations, generally categorized in either a central, trusted model (i.e., the query-response mode would fit here, because we are putting our trust in the entity handling the answering of a question—that entity knows the "true" answer and returns a sanitized answer) or a localized, untrusted model (i.e., where no "real" data is shared without undergoing a process of sanitization first). *See generally* Damien Desfontaines, *Local vs. Central Differential Privacy*, TED IS WRITING THINGS, https://desfontain.es/privacy/local-global-differential-privacy.html [https://perma.cc/2AG9-QY45] (Sept. 30, 2021). In this Article, we discuss primarily the query-response implementation of differential because it is somewhat easier to understand when learning the concept. Questions of "how private" are easily mapped, for example, to how accurate the response to a specific query is. Reasoning about how differential privacy applies to something like an entire dataset produced with differential privacy or how privacy is affected in the local version of differential privacy is out of the scope for this Article; however, the two-step test introduced here would nonetheless be applicable. *See, e.g.*, Xingxing Xiong et al., *A Comprehensive Survey on Local Differential Privacy*, SECURITY & COMM. NETWORKS, Oct. 8, 2020, art. no. 8829523, at 1, 1–3 (2020) (providing a summary of techniques for local differential privacy and also discussing centralized differential privacy); NINGHUI LI ET AL., DIFFERENTIAL PRIVACY: FROM THEORY TO PRACTICE 1–30 (Elisa Bertino & Ravi Sandhu eds., 2022).

[147] *See also infra* note 201 and accompanying text (discussing a reidentification).

[148] *See infra* Section IV.A. This is partly why differential privacy has faced criticism in overprotection; it necessarily operates in a world where two similar datasets exist, and reasons about what information is nonetheless learned from that world. *See infra* note 206 and accompanying text.

[149] *See id.* at 1–2 ("Tracing can be significant if, for example, the dataset comprises medical records of participants in a pharmaceutical trial or patient records from an abortion clinic.").

many of the standards in use today is that new attacks are later invented that undermine the assurance of outdated methods to sanitize data—what can you do with anonymized Massachusetts hospital information?[150] With this in mind, we turn to identifying an aspect of an ε-differentially private mechanism that is transferrable to a legal understanding of statutorily mandated data confidentiality.

2. *Legal Comparator*

Distilling a legal comparator from an ε-differentially private mechanism first requires an understanding of what various values of epsilon mean for a privacy loss. To aid this understanding, it is helpful to visualize the bounds of Algorithm 2's mechanism (as shown in Figure 4).[151]

That mechanism had an epsilon value of 1.098, which produced an upper bound of 75%. In the best-case scenario for an attacker, an observed output would be known to be real with a 75% confidence. Stated otherwise, if the attacker sees that an output to "are you a member of the Communist Party" is yes, then the attacker has a 75% confidence level that this was the participant's real answer—there is a 75% chance that this person is a member of the Communist Party. A visualization here allows us to more fully contextualize that 75%.

[150] *See* Ohm, *supra* note 11, at 1720–21 ("At the time [the Massachusetts Group Insurance Commission (GIC)] released the data, William Weld—then Governor of Massachusetts, assured the public that GIC had protected patient privacy by deleting identifiers. In response—then graduate student Sweeney started hunting for the Governor's hospital records in the GIC data. She knew that Governor Weld resided in Cambridge, Massachusetts, a city of fifty-four thousand residents and seven ZIP codes. For twenty dollars, she purchased the complete voter rolls from the city of Cambridge—a database containing, among other things, the name, address, ZIP code, birth date, and sex of every voter. By combining this data with the GIC records, Sweeney found Governor Weld with ease. Only six people in Cambridge shared his birth date; only three of them men, and of them, only he lived in his ZIP code.") (citing Sweeney, *supra* note 112). To be sure, this example relies on data which did not adhere to a statute like HIPAA. Nonetheless, this canary does exemplify the necessity for sanitization standards to be futureproof. *See also infra* notes 194–196 and accompanying text (applying these same types of reidentification attacks to statutorily-compliant datasets).

[151] *See supra* notes 134–140 and accompanying text (discussing where the probabilities come from).

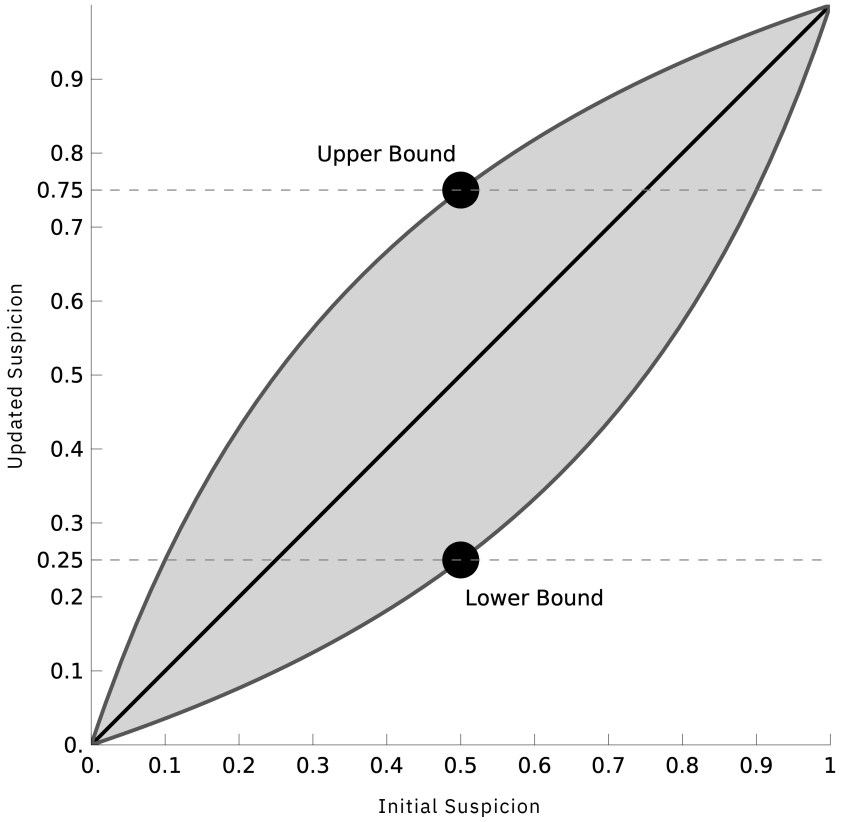

Figure 5. Bounding of The Randomized Response Algorithm[152]

The attacker in this mechanism has a 50% chance of correctly guessing an output a priori: an answer to a yes or no question is either yes or no. This knowledge is represented as the initial suspicion found along the *x* axis, at the 0.5 mark. Tracing this initial suspicion value vertically to the *y* axis will end at the thick black diagonal line, which represents what may be thought of as the home base position (i.e., the attacker did not learn anything from initial to updated suspicion). Using Figure 4's Algorithm 2, we found that the attacker's guess may be adjusted by at most 25% for a highest-possible confidence of 75%. This is represented by vertically adding 25% to that diagonal line, ending at 75%

[152] This image (showing the bounding of the randomized response algorithm) is a modified version of a figure found in Dr. Desfontaines's dissertation. *See* Damien Desfontaines, Lowering the Cost of Anonymization 1, 26–30 (2020) (Ph.D. dissertation, ETH Zurich), https://desfontain.es/thesis.pdf [https://perma.cc/352J-4SH2] (Fig. 2.1, epsilon set to the natural logarithm of three).

along the *y* axis. This point on the *y* axis is the a posteriori confidence, belief in the correctness of an output after seeing the mechanism's output.[153]

The final value here is known as the upper bound—an adversary can gain no more confidence when witnessing a mechanism's output than this percentage (i.e., that a provided answer by a mechanism is the real, truthful answer). The lower bound moves the confidence in the opposite direction and represents the best-case scenario for a respondent. Based on an observed output of a mechanism (i.e., a "no" answer in randomized response), the attacker may lose confidence in a guess (e.g., you thought there was a 50% chance of something happening, but when seeing a particular output value, your confidence drops to 25%).[154] Another way to think of these two boundaries is that not all answers are equal, some answers may be more likely than others, and therefore an attacker's confidence may change depending on the observed output. This change occurs because of how the mechanism is built.

To practicalize this dance of probabilities,[155] imagine you owned a crystal ball which tells you whether it will rain tomorrow: yes or no. Unfortunately, because the ball is magical, it is regulated, and you are only allowed to access predictions from the ball which have been sanitized using differential privacy. Further assume that you know the mechanism the crystal ball uses has an epsilon value of 1.098. Given that there is, at baseline, a 50% chance that it will rain tomorrow, if your crystal ball answers "yes," then you can be 75% confident that it will rain tomorrow—and this might be high enough for you to carry an umbrella. The output of the mechanism, even though differential privacy is being used, greatly impacted your decision to carry an umbrella.

On the other hand, assume the crystal ball were using a (.08)-differentially private mechanism to sanitize its future-predicting outputs.[156] If you had an a priori guess that it would rain tomorrow, 50%, and the crystal ball said "yes"—i.e., the same setup from before, with a revised epsilon value—then you would only have gained a 2% boost in confidence. You are now able to say there is a 52% chance of rain tomorrow—and that might *not* be high enough for you to take an umbrella. In other words, learning the output of this particular (.08)-differentially private mechanism does very little for your choice in umbrella encumbrance.

[153] While the initial suspicion's value depends on the type of question being asked (e.g., a binary yes or no question or an ordinal definitely, probably, equally likely question) the updated suspicion's value depends on epsilon. *See infra* Section III.C.

[154] This work focuses on the upper bound because it represents the best-case scenario for an attacker, allowing us to make more concrete statements which cover all possible outputs from a mechanism.

[155] These are loose terms because the random-age mechanism does not have the same epsilon value as Algorithm 2 nor does it have the same initial suspicion (i.e., there is not a binary "yes" or "no" answer, but rather a range of numbers that would be provided by the mechanism).

[156] *See also* Vito D'Orazio et al., Differential Privacy for Social Science Inference, Paper Presented at the Summer Meetings of the Society for Political Methodology (July 24, 2015), http://www.sas.rochester.edu/psc/polmeth/papers/Dorazio_Honaker_King.pdf [https://perma.cc/95MS-Y24L] (showing how various epsilon values affect the attacker's best-guess scenario: (100 * 50) / (50 + e^(-(ln(3)))(100-50)) = 75% as the updated suspicion value, assuming an initial suspicion of 50% and a epsilon value of ~1.098).

This fluidity in confidence is what must be translated into legal language. At a high level, lower epsilons mean that the data provided by a mechanism is more sanitized, and a statute that is highly sensitive to the risk of a privacy loss (i.e., risk of reidentification) would be more likely to approve the mechanism's outputs. However, there are a few important nuances not captured by such a cursory view. Three options may exist for the accurate and portable packaging of a mechanism's risk of reidentification.[157] Each of these options is discussed in turn.

a. Epsilon Alone

One possibility for translating a mechanism's legal risk is simply using epsilon alone. On the positive side, this approach places the focus on an easily adjustable quantity, allowing simple changes in epsilon to reposition the legal viewpoint of a mechanism's sanitization abilities. The downside, however, is that this approach is not very granular. Low epsilons may be considered more private, as "small [epsilons] are happy epsilons," but distinguishing between an epsilon of .01 versus .05 versus 1.0 would be practically difficult.[158] At the same time, this could impact a decision by a court given that not all data are created equal, and the purposes of data exploration are also not equal (i.e., some objectives are more worthwhile than others).[159] If a court has trouble distinguishing between "small" epsilons, then it could lead to permissible sharing when the risk is, in reality, too high.

Additionally, there is no context provided when considering an epsilon value by itself, which may produce a rubber-stamping effect on certain mechanisms. The quantity being assessed here should be the mechanism's ability to provide an attacker with a lot or a little information. Simply looking at epsilon alone does not provide a sense for how much information is being gained by the attacker. Indeed, an epsilon value of 1.098 may seem low, but comports with a 25% boost in confidence when observing some outputs. Depending on the particular scenario and an initial suspicion probability, a 25% boost could be an untenable amount of privacy loss. Therefore, epsilon alone is likely a nonideal fit for a legally portable understanding of differential privacy.

b. Upper Bounds

A second option for a legal comparator may be to consider the upper bound produced by a mechanism (e.g., the 75% in Algorithm 2). This approach has the benefit of capturing the worst-case scenario for any users' data that may be in

[157] True enough, there are likely many more options than this. We would urge more technical work to look into the guess difference and assess other viable options for identifying a proxy value for reidentification risk.

[158] Dwork et al., *supra* note 122, at 7–8; *see also* NEAR & ABUAH, *supra* note 130, at 19 ("How should we set ϵ to prevent bad outcomes in practice? Nobody knows. The general consensus is that ϵ should be around 1 or smaller, and values of ϵ above 10 probably don't do much to protect privacy—but this rule of thumb could turn out to be very conservative.").

[159] Dwork et al., *supra* note 122, at 8. For example, a social media company exporting all of its data to learn how to best increase click counts is an in-kind difference to a hospital releasing health records to help patients who suffer from a particular medical illness.

the dataset. As not all answers provided by a mechanism carry the same amount of risk (e.g., in the randomized response mechanism discussed in Section III.A above, observing a "yes" answer carried the most risk, with an upper bound of 75%), this quantity appropriately captures all possible output, best case and worst case for the attacker.

The downside to this approach, however, is that *only* the upper limit is taken into consideration. In this way, this measurement may oversell the adversary, leading to a court being more wary of a situation that presents less risk than perceived. For example, at an initial suspicion level of 75% and an epsilon value of one, the attacker ends with an 89.08% upper bound percentage.[160] Although 75% is fairly high to begin with, the epsilon value being used here is in some ways low. Despite this, a nearly 90% upper bound probability is unlikely to be approved by a court looking to protect a user's data.

In summary, regardless of how it may be beneficial to consider the worst-case scenario given that we would be matching this number with the maximum risk permitted by a statute, this comparator ignores important context like a priori guessing ability, which provides useful context for a court to consider. For this reason, the upper bounds are less likely to be the best fit for the type of legal comparator we are looking for.

c. Guess Difference

A final possibility is to use what we deem the guess difference. The guess difference is the difference between the initial suspicion and the upper bound; in short, taking out what the attacker already knew and only keeping what was learned from the algorithm's output in the best-case scenario for the attacker. For example, an initial suspicion of 50% with epsilon 1.098 (i.e., Algorithm 2) produces an upper bound of 75%; therefore, we have a guess difference of 25%, the difference between the initial suspicion and the upper bound.

This approach allows us to take into consideration the fact that some questions are more privacy sensitive than others by relying on the upper bound, but tempers this by removing the default guessability of a query. To be sure, in this way, the guess difference may undersell the attacker's overall guessing ability. For instance, it might seem odd that a high initial suspicion and low epsilon value nonetheless produces a low guess difference score, despite the fact that the attacker had a high likelihood of guessing initially and that guess was only made stronger after seeing the output of a mechanism.[161] Looking closely at the aims of differential privacy, however, shows that this is likely a moot point.

Differential privacy does not concern itself with information not gleaned via the dataset.[162] Imagine that an individual who has a particular disease par-

[160] *See* D'Orazio et al., *supra* note 156, at 7.

[161] It would seem that differential privacy is not protecting the output given the high initial guessing ability of the attacker, and the guess difference score would obscure this understanding because it produced a low value.

[162] This is an adaptation of the example provided by Dwork and Roth. *See* Dwork & Roth, *supra* note 1, at 215–16 ("A medical database may teach us that smoking causes cancer, affecting an insurance company's view of a smoker's long-term medical costs. Has the smoker been harmed by the analysis? Perhaps—his insurance premiums may rise, if the insurer knows he smokes. He

ticipates in an experimental drug study where the data from the study is protected using differential privacy. Further imagine that the published results of the study are that the experimental drug increased life expectancy rates by one year. Would we say that differential privacy failed to protect this individual if the individual's insurance rates are increased after the insurance provider learns of this exact study and its conclusion? No.

The insurance company learned from the broad result published by the study, which differential privacy does not claim to protect. If, on the other hand, the insurance provider increased the individual's rates after querying the dataset and coming up with some confidence level that the individual was "in" this dataset, meaning the individual had the potentially life-threating disease, then we would say that differential privacy failed to protect the individual. Differential privacy allows us to draw hard lines around how much the insurance company may learn from the data—and guess difference captures that ability. For instance, we may say that the insurance company will never be able to increase a blind guess likelihood by more than 2%; a blind guess that this individual is a smoker cannot be confirmed by querying the data because the likelihood that that guess is correct will never be increased by more than 2%, no matter what result is found in the dataset. Stated otherwise, it would be illogical to conclude, based on the results of any query on this dataset—which the individual is in fact "in"—that the individual's rates should be increased. The insurance company may nonetheless increase the individual's rates, but would not be basing this decision on a reliable fact learned from the dataset.

Overall, the guess difference approach provides a singular, but context-filled legal comparator. This quantity highlights differences in risk when epsilon is small, allowing a court to meaningfully interpret the .01 to .05 to 1.0 epsilon range, it incorporates the worst-case scenario for any user who is in a dataset, by working with the upper bound set by a particular epsilon value, and it accords with preexisting considerations of reidentification risk, as discussed further in Section IV.A below. Therefore, we conclude that out of the three options discussed above, guess difference should be the quantity used to interpret the sanitization abilities of an ε-differentially private mechanism from a legal vantage. The following Section generalizes the guess difference as a proxy for a mechanism's risk of reidentification.

C. Step One: Reidentification Risk vis-à-vis the Guess Difference

Taking these options together leads to the conclusion that guess difference is the most appropriate legal comparator—guess difference may be considered a proxy value for the reidentification risk a mechanism encumbers. This option adequately balances the attacker's best-case scenario, but tempers that confi-

may also be helped—learning of his health risks, he enters a smoking cessation program. Has the smoker's privacy been compromised? It is certainly the case that more is known about him after the study than was known before, but was his information 'leaked'? Differential privacy will take the view that it was not, with the rationale that the impact on the smoker is the *same independent of whether or not he was in the study*. It is the *conclusions reached* in the study that affect the smoker, not his presence or absence in the data set.").

dence with the a priori guessability of the query. In this way, the measurement does not oversell or undersell the sanitization abilities of a mechanism. This metric will therefore form step one of our two-step test permitting the comparison between what differential privacy provides and data-protecting regulation mandates.

1. *Epsilon Visualized*

With that in mind, we may visualize a range of popular epsilon values in terms of the *guess difference* each mechanism provides:

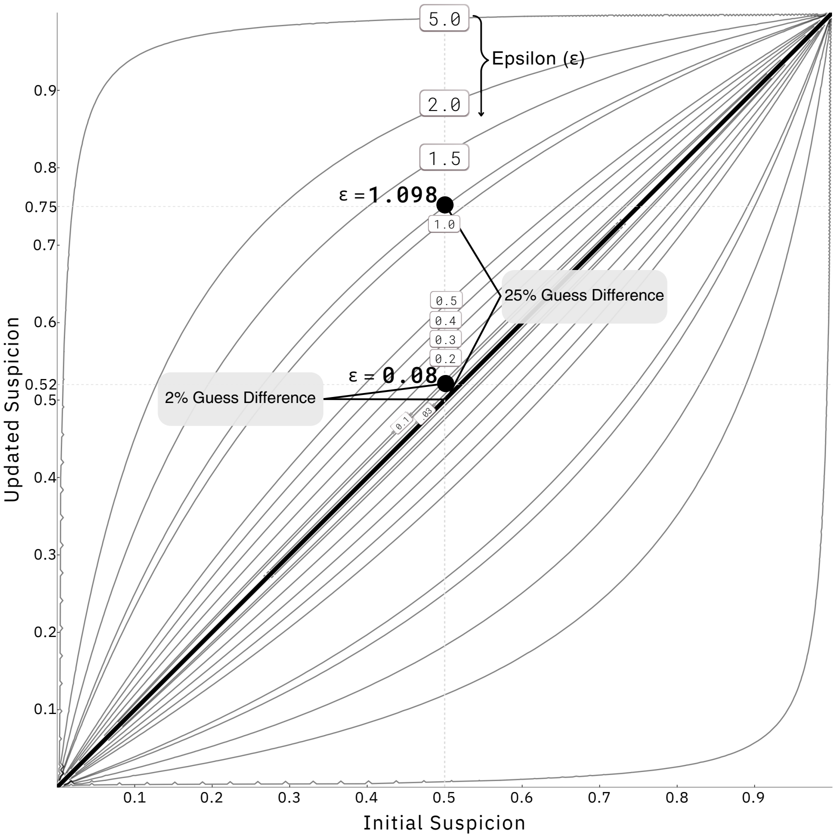

Figure 6. Guess Difference Visualized[163]

[163] This image is a modified version of a figure found in Dr. Desfontaines's dissertation. *See* Desfontaines, *supra* note 152 (adapted from fig. 2.2). An epsilon value of .08 has a 2% guess difference assuming an initial suspicion of 50%. This is found by taking the initial suspicion (assumed here at .5) and subtracting it from the updated suspicion (equated with epsilon set at .08, for a value of .52), in the appropriate order. The result is .52-.5=.02. Likewise, an epsilon value of 1.098 has a 25% guess difference. This is found by taking the initial suspicion (assumed here at .5) and subtracting it from the updated suspicion (equated with epsilon set at 1.098, for a value of .75). The result is .75-.5=.25. This simple analysis may be applied to any epsilon value as long as the initial

Figure 6 shows epsilon values (i.e., the curved lines) ranged .03 to 5, with smaller epsilon values found closer the thick black diagonal line. The diagonal line, as discussed in Section III.B.2 above, may be thought of as the home base for an inquiry when visualizing a mechanism this way (i.e., nothing is learned from initial suspicion to updated suspicion—perfect privacy[164]).

To find a guess difference using Figure 6,[165] first locate an initial suspicion value provided by a mechanism along the *x* axis (i.e., along the bottom).[166] Then, take note of the epsilon value of the mechanism under consideration. Each epsilon value is associated with a resulting line drawn across the top half and bottom half of the figure (upper and lower bound, respectively). This line may be called the epsilon line.[167] Trace the initial suspicion value along the *x* axis vertically until you hit the thick black diagonal line. Take note of this point (what may be called home base) and then keep tracing it until you cross the epsilon line. The point where the epsilon line meets the vertical line drawn by the initial suspicion value is called updated suspicion. Guess difference is equated by subtracting the update suspicion number from the home base position (i.e., where the initial suspicion line meets the thick black horizontal line). As an example, in Figure 6, we can see that the initial suspicion of .5 meets the epsilon line at .52 when using an epsilon value of .08. This would be a guess difference of .02 or 2%.

Overall, one can visually see the guess difference by looking at the distance between thick black horizontal line in that home base position and then measuring to the updated suspicion value. As a whole, this visualization allows us to see how larger values of epsilon affect the guess difference, with an epsilon value of 1.098 being much farther from the diagonal line than the .03 or .08 epsilon values.

and updated suspicion values are also known. *See also* D'Orazio et al., *supra* note 156, at 5–8 (same, providing a table of upper bounds for a sampling of epsilon and initial suspicion values).

[164] This illustration also allows us to easily see that higher epsilon values are "less private"—privacy degrades as we move away from the thick black diagonal line.

[165] This conceptualization of guess difference is intuitive for a Bayesian analysis of randomized response, but other scenarios using differentially private mechanisms exist, such as aggregate statistics. For more detail, see D'Orazio et al., *supra* note 156, at 5–8 ("Table 1 shows the effect of different epsilon values on our belief that [a provided answer is the "real" answer (i.e., T = 1)]. The left column is our prior belief that T = 1. Each column to the right contains an upper bound on our updated belief having learned [the output of the mechanism]. For example, if there is a 99% chance of John's political affiliation being known, and then we learn [the mechanism's output] with an epsilon of 0.5, then our belief about John's political affiliation can become at most 99.39%."); *see also supra* note 146 and accompanying text.

[166] This percentage is given based on the particular issue being solved with the differentially private algorithm.

[167] Flatter epsilon lines denote lower epsilon values while more curved lines denote higher epsilon values. This occurs because lower epsilon values do not provide as much of a boost in confidence as higher epsilon values do.

2. *Takeaways*

Assessing an ε-differentially private mechanism from a legal vantage may be easily accomplished by considering the guess difference—what may be deemed the risk of reidentification a mechanism accommodates. This value is found by knowing: (1) the epsilon value associated with the mechanism; (2) the initial suspicion value provided by a mechanism (i.e., likelihood of guessing a "real" output without seeing a mechanism's output); and (3) the updated suspicion value of the upper bound of a mechanism. When subtracting out the initial suspicion from the updated suspicion, derived using the updated suspicion and epsilon value, we arrive at the guess difference; essentially, the risk of reidentification a mechanism permits. From an attacker's perspective, we can guarantee that there is no more than a *guess difference* chance that an attacker will be able to take the output of a mechanism and say: "This is the real answer."[168]

Step one adds context to a mechanism and provides a legally framed benchmark that may be measured against to a variety of statutes to assess whether the mechanism produces private-enough data to permit sharing. The next Part introduces the legal corollary against which the guess difference is measured: a statute's threshold for reidentification risk.

IV. STEP TWO

The following Part examines step two: a statute's maximum allowance for reidentification risk. Step two, practically, requires a statute-by-statute inspection which is in many ways lackluster when attempted from the armchair. That said, an argument about why the risk of reidentification is at the heart of all data protective statutes (i.e., the applicability of step two), and why the quantity discussed in step one speaks the same language as step two will be necessary. Following these two arguments, we look at how HIPAA may be interpreted under the two-step test.

A. Statutory Privacy

Wearing a legal hat while considering the implications of differential privacy gives rise to two primary obstacles.[169] First, statutes regulating data do not speak in terms of a measurable privacy loss. Instead, shareable data is protected under explicate terms like "remove n identifiers" or ambiguous terms like "remove *any* information which could lead to identification." Regardless of the phrasing, however, both terms belie what sits at bottom: protection against the risk of reidentification. Second, when a statute does find itself associated with a measurable reidentification risk, one which sets the bar for permissible data sharing, the end result has been, in many ways, meaningless—the permissible

[168] We discuss the two-step test in regard to the query-type interaction with a differentially private mechanism, although other forms of differential privacy exist. *See supra* note 146 and accompanying text.

[169] For an overview of differential privacy and policy issues, see generally Vitaly Feldman et al., *Differential Privacy: Issues for Policymakers*, SIMONS INST. (June 29, 2020), https://simons.berkeley.edu/news/differential-privacy-issues-policymakers [https://perma.cc/Y76S-GZMH].

risk changes depending on the question being asked and the invention of novel, adversarial techniques, not to mention how these approaches are difficult to apply across a variety of statutes. For these reasons, before illustrating how our two-step test would stack up against a statute, it is necessary to: (1) illustrate how the risk of reidentification is at the heart of all statutes built to protect data; and (2) evidence how and why current measurements of reidentification risk fall short.

1. *The Heart of Statutory Privacy*

When drafting a statute intended to protect data, a common approach is to hinge that protection on the definition of PII.[170] It is *impermissible* to share PII, but *permissible* to share non-PII. VPPA's prohibition on sharing "information which identifies a person" or COPPA's prohibition on sharing "individually identifiable information about an individual collected online" are par for the course. This is true even for regulations which seem to swallow any and all data—for example, the GDPR.[171]

The difficulty with sharing data while trying to comply with these regulations, however, is that it creates a red herring, a "find-the-gaps" exercise that obfuscates the intent of the regulation. The exercise plays out like this[172]: it is permissible to share data as long as the data does not include a specific set of attributes that could, would, or do link to an identity[173] or it is permissible to share data as long as the actors (i.e., a specific type of entity[174] which is regulated, as opposed to an unregulated entity) or substance (i.e., a specific type of data which is regulated, as opposed to unregulated data[175]) are not subject to the law. Unfortunately, this exercise provides seemingly simple answers (i.e., look for the gaps when trying to share protected data) which break when considering far reaching statutes like the GDPR.

[170] This is true at least when it comes to regulation in the United States. Countries outside of the United States use a variety of other terms, such as "personal data." *See generally* W. Gregory Voss & Kimberly A. Houser, Pers*onal Data and the GDPR: Providing a Competitive Advantage for U.S. Companies*, 56 AM. BUS. L.J. 287, 292 (2019) (discussing PII and "personal data").

[171] *See* Regulation (EU) 2016/679 of the European Parliament and the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1 [hereinafter GDPR].

[172] *See generally* Stacey A. Tovino, *Not so Private*, 71 DUKE L.J. 985, 990–91, 1000–19 (2022) (focusing on health data).

[173] *See* COPPA, Children's Online Privacy Protection Act, 15 U.S.C. § 6501(8) (2017); *see also* C.F.R § 164.514 (2018) (HIPAA safe harbor stripping provision).

[174] *See* FERPA, Family Educational Rights and Privacy, 34 C.F.R. §§ 99.3, 99.1(a) (2014) ("[T]his part applies to an educational agency or institution to which funds have been made available under any program administered by the Secretary if (1) [t]he educational institution provides educational services or instruction, or both, to students; or (2) [t]he educational agency is authorized to direct and control public elementary or secondary, or postsecondary educational institutions.").

[175] *See id.*; *see also* Elana Zeide, *Student Privacy Principles for the Age of Big Data: Moving Beyond FERPA and FIPPS*, 8 DREXEL L. REV. 339, 359 n. 102 (2016) (online corrected) ("A significant amount of potentially sensitive student information falls outside the statute's protection due to narrow definitions of what constitutes PII maintained in a student's education record and an exclusion for institutionally-defined 'directory information.'").

The GDPR's reach on regulated data is one of the broadest, swallowing any data "relating to an identified or identifiable natural person."[176] This means that, absent statutorily prescribed exceptions, data may not be shared.[177] Even pseudonymized data (i.e., data which has undergone privacy-protective measures, but which may nonetheless be joined with auxiliary information and lead to the identification of a person) is unshareable without statutory proscriptions like consent and minimization. The regulatory line does stop, however, at anonymized data: "The principles of data protection should therefore not apply to anonymous information [i.e.,] . . . personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable."[178] This statement highlights the core issue: How much sanitization is enough; when is data *anonymized* to the point where the individuals it describes are no longer identifiable? In some light, the GDPR's anonymization requirement may seem impossible—100% anonymization would require 0% utility in certain use cases.[179]

That the GDPR's reach stops short of anonymized data highlights the impetus behind finding the gaps—the gaps highlight permissibly shareable areas because these are areas the drafters felt needed little or no privacy protection. These are areas where there is little or no risk of a privacy loss, a reidentification. Stated otherwise, regulating the risk of privacy loss sits on a spectrum. Along this spectrum are those sanitization techniques which produce mitigated, but not eliminated, privacy loss (i.e., pseudonymized), and also those techniques which reduce the risk of privacy loss to the point where it is merely theoretical (i.e., anonymization).[180] To be sure, no method of data release is completely risk free,[181] even differential privacy. Nonetheless, all data protective statutes do reg-

[176] *See* GDPR, *supra* note 171, art. 4(1). Uncommonly for a data protective statute, the GDPR considers inferences as regulated data, inferences *which may not even be correct*. *See id.*

[177] If all or most data arguably relates to an identifiable person (i.e., the database of ruin) then there are few if any carveouts for unregulated data. *See supra* note 53 and accompanying text.

[178] *See* GDPR, *supra* note 171, Recital 26.

[179] *See* Morgan Lewis, *The eData Guide to GDPR: Anonymization and Pseudonymization Under the GDPR*, JDSUPRA (Dec. 9, 2019), https://www.jdsupra.com/legalnews/the-edata-guide-to-gdpr-anonymization-95239/#_ftn14 [https://perma.cc/QKC3-NF39] (discussing cases concerning the line between anonymization and pseudonymization).

[180] Indeed, if data is useful at all, then it has a nonzero risk privacy loss when released. *See* DATA PROTECTION COMM'N, GUIDANCE NOTE: GUIDANCE ON ANONYMISATION AND PSEUDONYMISATION 1, 7 (June 2019), https://www.dataprotection.ie/sites/default/files/uploads/2019-06/190614%20Anonymisation%20and%20Pseudonymisation.pdf [https://perma.cc/Z9MB-9VXT] ("It is not possible to say with certainty that an individual will never be identified from a dataset which has been subjected to an anonymisation process."); *see also id.* at 3 ("There is a lot of research currently underway in the area of anonymisation, and knowledge about the effectiveness of various anonymisation techniques is constantly changing. It is therefore impossible to say that a particular technique will be 100% effective in protecting the identity of data subjects . . . ."); Lewis, *supra* note 179, (discussing how GDPR operates on the difference between "anonymous" data and "pseudonymous" data). That said, the risk is to such a degree that it becomes minutia.

[181] This is formally known as the Fundamental Law of Information Recovery. *See, e.g.*, Dwork et al., *supra* note 122, at 2 ("A compendium of results, colloquially known as the Fundamental Law of Information Recovery, tells us that overly accurate answers to too many questions can destroy any reasonable notion of privacy."); Dinur & Nissim, *supra* note 118, at 204 ("We show that any database algorithm that is within o(√n) perturbation, is non private with respect to polynomial time adversaries. More accurately, we show that whenever the perturbation is smaller than √n,

ulate risk, the "risk" (i.e., the spectrum) of identifying individuals within a dataset. PII is merely a proxy describing that reidentification risk—so to, as we introduce, is the guess difference, albeit a technical understanding of that risk.

A good way to measure this spectrum is to waypost the ease at which reidentification occurs, the point at which anyone can point at a record and say "I know who that is."[182] Raw data would be at one end of the spectrum[183] and anonymized data would be at the other, with pseudonymized data and data with n stripped identifiers (e.g., HIPAA) lying closer to anonymized data. That an individual has been identified, regardless of where this occurs along the spectrum, means there is no additional loss that might occur; it is merely the question of how likely released data could reach that point, and this is what statutes regulate—where on the spectrum the to-be-released data must fall.

This theme is not without support from the courts. In *Pub. Citizen Health Research Grp. v. Pizzella*, when discussing plaintiff's argument that an OSHA requirement that was more privacy protective than its incumbent was not sufficiently justified by the record, the court found ample support for the new regulations because the risk of reidentification under the previous requirements was too high.[184] In other words, OSHA's position on requiring further privacy protections was reasonable because of the ultimate harm OSHA sought to protect against, reidentification.[185]

Furthermore, several courts focus directly on reidentification risk when considering what a statute requires to release protected—but sanitized—data. Partially, this comes from HIPAA's statutory language that strikes very close to the risk of reidentification (e.g., "the risk is very small that the information could be used . . . to identify an individual"[186]), but courts have also come to this conclusion on their own. In *Sander v. Superior Court*, when discussing whether records could be released pursuant to a FOIA-themed statute, the court made its determination in large part based on the risk of reidentification that a release would incur; in *Setinberg v. CVS*, the court, when providing guidance on whether the sharing of deidentified, but possibly reidentified, records would be permissible under HIPAA, found that assessment by an expert about the reidentification risk the released records bore would be necessary; and in *Cohan*

a polynomial number of queries can be used to efficiently reconstruct a 'good' approximation of the entire database."). In more simple terms, what Dinur and Nissim found was that revealing anything useful about a dataset, on an ongoing basis, will eventually destroy any privacy that may have been found in the dataset in the first place—and the point at which all privacy is destroyed occurs sooner than you might think. *Id.*

[182] *See also supra* Section III.B.1 (describing reidentification risk in terms of step one).

[183] A caveat is needed here given that not all data is easily attributable to a person.

[184] *See, e.g.*, Pub. Citizen Health Rsch. Grp. v. Pizzella, 513 F. Supp. 3d 10, 25–26 (D.D.C. Jan. 11, 2021).

[185] *See id.* ("OSHA determined that 'even if PII could be completely removed from the data, concerns about re-identification would remain.' Moreover, the Revised Rule states that 'particularly in a small town,' information like 'what was the employee doing just before the incident occurred, what happened, and what was the injury or illness' could allow re-identification." (citing F84 Fed. Reg. 384)).

[186] *See* C.F.R § 164.514.

v. Ayabe, the court interpreted HIPAA's expert-deemed safe harbor to rest on a determination that the reidentification risk was "very small."[187]

In summary, statutes protecting data are foundationally regulating the risk of reidentification. Statutes may go about this task with a variety of artisanal linguistic options, but the core of what is being regulated is privacy loss, which may be quantifiably expressed as the risk or likelihood of reidentification.[188] The more privacy sensitive a statute, the less risk is tolerated; the less privacy sensitive a statute is, the more risk is tolerated. The next question, therefore, concerns the permissible level of risk, quantitively, that a statute allows. Here, unfortunately, despite seeming clarity, we find an unworkable standard.

2. *Moving Targets*

Statutes like HIPAA, which have been subjected to a fair amount of technical interpretation regarding whether data is "sanitized enough" to meet the statute's reidentification risk threshold, have fallen into a rut when it comes to defining permissible reidentification risk. The crux of this rut centers on how the technical literature has acquiesced to a definition of reidentification that was stated loosely at first, but which has, over time, grown to take on a meaning of its own. In turn, this definition has worked its way into the courts as fact. The incorrect statement looks like this: HIPAA allows for data to be released if there is a .04% to 25% risk of reidentification.[189]

To begin, only two ways exist to release regulated data under HIPAA. The first option is for a data steward to strip the record of a series of explicit attrib-

[187] *See* Sander v. Superior Ct., 26 Cal. App. 5th 651, 660, 237 Cal. Rptr. 3d 276, 283, 2018 Cal. App. LEXIS 755, at *13, 2018 WL 4024906 (Cal. App. 1st Dist. Aug. 23, 2018); Steinberg v. CVS Caremark Corp., 899 F. Supp. 2d 331, 335–37 (E.D. Pa. 2012); Cohan v. Ayabe, 132 Haw. 408, 424, 322 P.3d 948, 964, 2014 Haw. LEXIS 95, at *52–53, 2014 WL 783132 (Haw. Feb. 27, 2014).

[188] For a discussion of how a reidentification may come about, see Michelle N. Meyer, *Reflections of a Re-Identification Target, Part I: Some Information Doesn't Want To Be Free*, FAC. LOUNGE (May 24, 2013), https://www.thefacultylounge.org/2013/05/reflections-of-a-re-identification-target-part-i-some-information-doesnt-want-to-be-free.html [https://perma.cc/N7U9-4VCK] ("I wanted to donate my genotype and phenotype data to a project committed to open access research . . . . With a little digging, I learned that there had in fact been two re-identification demonstrations involving the PGP. . . . Latanya Sweeney used the algorithm based on zip code, birth date, and gender that she made famous in her 1997 re-identification of Massachusetts Governor Bill Weld—and also read some participants' names directly from their decompressed 23andMe files. And Yaniv Erlich used the algorithm based on Y-chromosome data and surnames that he published in Science earlier this year. . . . After a little more digging, I found the paper in which Latanya reported her algorithm. Once I had the chance to read about it and to compare it to the information I had provided in my PGP profile page, I concluded with about 99.9% certainty that I was not among those who had been reidentified by either attack: I had not provided all of the information used in Latanya's algorithm. *Heeding the PGP's own pop-up warning, I had scrubbed my 23andMe file of my name before uploading it* (although I leave room for a 0.1% chance that I somehow did not thoroughly scrub it).") (emphasis added).

[189] No, 25% is not a typo. The range is embarrassingly large, as noted in notes 194–196, because the measurement of how to calculate that risk keeps changing, along with the datasets underlying each calculation. *See, e.g.*, Sweeney et al., *supra* note 92; *see also* Tovino, *supra* note 172, at 996 (discussing reidentification rates in the 3–10% range).

utes like name, email address, and social security number.[190] The second is to rely on an expert to certify 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."[191] To be sure, neither of these options mentions a number between .01 to 25; yet, a short survey on what "very small" means to a statistical audience would draw that conclusion: "Based on the nationally accepted standard of re-identification risk no greater than . . . 0.04"[192] and "[w]hen the redacted data contained the exact birth year, as allowed by HIPAA Safe Harbor, we correctly identified [25 percent in a subsection of the dataset] . . . . In comparison, earlier studies found unique re-identification rates in data that adhered to the level prescribed by HIPAA Safe Harbor to be much lower, namely 0.013 percent and 0.04 percent."[193] In these examples, researchers are taking it as a fact that HIPAA requires some level of reidentification risk to permissibly share data. This alone is not a problem, and accords with the two-step approach we argue for in this Article. The problem occurs, however, because of how that reidentification risk is being calculated—the number keeps changing.

These percentages come from attempts to read hard numbers into HIPAA, but with non-futureproof methods. The original idea was to scrub records pursuant to HIPAA's safe harbor provision (i.e., remove the series of n identifiers), check how many resulting records were nonetheless unique, and then report that number as the *inferred* reidentification risk. The argument would go like this: this data release is permissible because it is the same level of sanitization that is accepted under HIPAA's explicit safe harbor provision.

Professor Sweeney debuted this method over two decades ago, finding that "0.04% . . . of the population of the United States is likely to be uniquely identified by values of {gender, year of birth, ZIP}."[194] Since then, a line of work has sprung up reapplying the approach—but the numbers have changed drastically over time, ranging anywhere from 0.01%[195] to 25%.[196] The reason for this moving target is owed to a property of differential privacy that is not found in assessments of uniqueness-based metrics like the .04% rule—futureproof.

[190] The stripping of 17 identifiers (plus a catchall) is the second. C.F.R § 164.514(b)(2).

[191] *Id.* (emphasis added). To be sure, the most promising option is the second, in that it is flexible to the times (i.e., a sledge hammer of deidentification may not be needed, though this is the only option for the second safe harbor) and adequately assures privacy (i.e., a "person with appropriate knowledge" must make this determination). *See id.* § (b).

[192] Victor Janmey & Peter L. Elkin, *Re-Identification Risk in HIPAA De-Identified Datasets: The MVA Attack*, *in* AMIA ANNUAL SYMPOSIUM PROCEEDINGS 1329, 1329 (2018).

[193] Sweeney et al., *supra* note 92, at 2.

[194] Latanya Sweeney, *Simple Demographics Often Identify People Uniquely* (Carnegie Mellon Univ., Data Privacy Working Paper No. 3, 2000), https://dataprivacylab.org/projects/identifiability/paper1.pdf [https://perma.cc/Y8G5-7K7Q].

[195] *See* Kathleen Benitez & Bradley Malin, *Evaluating Re-Identification Risks with Respect to the HIPAA Privacy Rule*, 17 J. AM. MED. INFO.. ASSO'N 169, 169 (2009) ("The percentage of a state's population estimated to be vulnerable to unique re-identification (i.e., g = 1) when protected via Safe Harbor and Limited Datasets ranges from 0.01% to 0.25% . . . .").

[196] *See* Sweeney et al., *supra* note 92, at 51 ("The number of correct re-identifications found in the HIPAA Safe Harbor–compliant data having exact year of birth is remarkable (25 percent uniquely and correctly re-identified by name).").

Differential privacy, in assuming the worst-case scenario for privacy, considers all information that is currently available or may be available in the future. This is why the definition of differential privacy focuses on the two versions of the mechanism's input which differ in a small way. This is also why differential privacy, according to some, is like using a sledgehammer to crack a nut (i.e., a great loss of utility at the gain of privacy).

Yes, differential privacy is in some ways excessive, but this property also allows differential privacy to make guarantees in perpetuity. The .04% rule, on the other hand, only looks to the narrow situation of a particular identifier-stripping provision and a particular (*unique/total*) equation. Using vague definitions "unique" and "total" means the resulting calculation will oscillate as definitions change over time. Moreover, this equation was written for HIPAA as an interpretation of the identifier-stripping provision; statutes which lack such a provision would require additional inferential leaps to make the argument that the equation generalizes.

Ultimately, however, despite the reasons why this approach may be ill-suited for statutory interpretation and why better solutions might exist, the approach is nonetheless making its way into the judicial branch. As seen in the *Sanders* case, one of the only cases to debate methods of data sanitization in terms of what level of sanitization is required by a statute, the court held in dicta that any proposed method of sanitization, as offered by the Plaintiff to access Bar admissions data, warranted too high a risk of reidentification—when measured against the HIPAA standard of ".02% to .22%."[197] Though not precedential, the reasoning is persuasive, suggesting that other courts may follow suit in trying to apply a unique-record-count fraction as a proxy for risk of reidentification.

This is wrong. What should a court do if tomorrow the target is moved from .02% to 2%? Yesterday's sanitization method may have failed, but *ex post facto*, it succeeds; the meaning of the risk of reidentification, as a result, is diluted. A better definition of risk of reidentification is needed; differential privacy's "guess difference" is needed.

In summary, current methods of analyzing the risk of reidentification point in the right direction, but fall short. A better approach is to rely on differential privacy for its futureproof property. In this way, the legal comparator introduced in Section III.C will hold despite any number of auxiliary pieces of information which come to the fore, and despite new attacks which attempt to pierce privacy's veil.

B. Step Two: A Statute's Measurement

Given that (1) all privacy-protective statutes, at bottom, aim to regulate the risk of reidentification; and (2) the risk of reidentification may be understood as the "guess difference" value, the next question to ask is: What amount of reidentification risk does a particular statute permit? Due to statutory diversity,

[197] Sander v. State Bar of California, No.CPF-08-508880, *21–22 (Nov. 7, 2016) (on file with author).

it would be shallow to argue for a definitive and generalizable answer in an Article like this. Instead, the Article provides a hypothetical revolving around HIPAA, discussing what *likely* risk the statute tolerates and what *likely* settings of epsilon would meet that risk threshold.

Consider a set of publicly accessible hospital records and a sample query: "How many individuals in this dataset have Crohn's disease." Further imagine that the interaction with this dataset is filtered through a differentially private algorithm (i.e., the query mode of differential privacy[198]), hence the reason it is publicly accessible. To be sure, raw data is transformed into noisy data before its receipt by the individual making the query.

For concreteness, the below table visualizes this information, both with a real answer of one and a real answer of 5,000; either one individual in the dataset has Crohn's disease or 5,000 do (i.e., consider this the ground truth). Both epsilon values, .08 and 1.098 (see Figure 6, visualizing epsilon values and guess difference), are compared across a sampling of ten possible answers a mechanism might provide, with averages noted in the last row. The data has been "post processed" by rounding to positive, whole numbers, and the Laplace method was used to generate noise.[199]

	Real Answer: 1		*Real Answer: 5,000*	
	$\varepsilon = 0.08$	**$\varepsilon = 1.098$**	**$\varepsilon = 0.08$**	**$\varepsilon = 1.098$**
1:	23	1	5,014	5,000
2:	3	2	5,003	5,000
3:	8	1	4,994	5,001
4:	12	1	5,020	4,999
5:	2	2	4,970	4,999
6:	1	2	5,000	5,000
7:	2	1	4,983	5,000
8:	9	1	4,996	4,998
9:	12	0	4,993	5,000
10:	3	1	5,005	5,000
Average	**4**	**1**	**4,998**	**5,000**

Table 4. Epsilon Affecting Differentially Private Queries[200]

Would a mechanism using an epsilon value of .08, assuming this particular data setup, and answering this particular question, run afoul of HIPAA?

[198] There are many different ways to use differential privacy, but one of the easiest, didactically, comes from thinking of its use in the query-response mode: you ask a question and a differentially private mechanism returns a sanitized response. *See supra* note 146 and accompanying text.

[199] For a further explanation of how these particular numbers may have been derived, *see infra* note 200 and accompanying text.

[200] The code, which built a "toy" (.08)-differentially private mechanism, is simple to code, using the 'diffprivlib' Python library, see Holohan et al., *supra* note 17:

```
from diffprivlib.mechanisms import Laplace # IBM's differential privacy library
laplace = Laplace(epsilon=.08, sensitivity=1) # alternatively epsilon=1.098
print(laplace.randomise(1)) # alternatively laplace.randomise(5000)
```

Pursuant to our two-step test, we first ask what is the risk of reidentification (i.e., the guess difference) a (1.098)-differentially private mechanism affords?[201] The answer is 25%, according to Section III.B.2.[202] With this in mind, we turn to step two, the maximum risk of reidentification a statute permits. Because HIPAA would apply here, we may look to the "very small" language found in the statute regarding expert-deemed "safe" data release. Is "very small" a term typically associated with 25%? If there was a 1/4th chance of rain tomorrow, would it be reasonable to call that a "very small" chance? In the balancing act a court would engage in, a one fourth chance that an individual is reidentified is likely too high for HIPAA. Therefore, this mechanism, with this epsilon value, under this statute, would likely not produce legally compliant outputs.

If, on the other hand, the epsilon value was .08, would this change the outcome? In this case, the first step, the guess difference of the (.08)-differentially private mechanism, is 2%. An attacker would gain a mere 2% increase in confidence that a provided answer is the real answer; from something like a 50% chance that an output is truthful to a 52% chance. The second step would be inquiring whether HIPAA's maximum permitted risk of reidentification is less than or equal to 2%. Is this setup *likely* to be permitted by HIPAA? Yes. Although a court would have to balance the competing interests and risks being presented, a 2% chance of reidentification—especially when HIPAA does not require a 100% free-from-all-harm guarantee—is likely sufficient. A few points from this short hypothetical are notable.

For starters, this permittance would cover a .08 epsilon value and lower. What this means is that a data steward would be able to interpret a stamp of approval on .08 to also mean that .07, .05, or .01 epsilon values are all appropriate. This provides freedom to adjust mechanisms to suit individual use cases while maintaining compliance.[203]

Secondly, it is notable that *four*, the average of the answers provided in the .08 epsilon column in Table 4, is fairly removed from the real answer of *one*. In fact, many of the responses in the .08 epsilon column appear to be inaccurate, though very privacy preserving. In short, this occurs because our numbers are

[201] A large caveat here is equating a known-truthful answer with a reidentification. That said, this practical scenario nonetheless communicates the quantity (i.e., reidentification risk) we are trying to capture with the guess difference proxy. In a more archetypal, yet less realistic case, the attacker would know that an individual, Alice, is in the dataset, know that everyone except Alice does not have Crohn's disease, but not know Alice's status. A reidentification would occur when a link is made between Alice and Alice's data—the output of the mechanism is confirmed to be true. With our guess difference proxy, we know that the output of the query may give a boost to the attacker's confidence level up to some threshold. For example, using an epsilon value of 1.098 we have a 25% boost (i.e., upper bound) in confidence; the attacker, in this less-realistic example, is now 75% confident that Alice has Crohn's disease. And that confidence is likely to offend HIPAA's "very low" language, there is not a very low chance of a reidentification occurring. Table 4 illustrates how, when there is only one individual being considered (i.e., a "real" answer of one), a 1.098 epsilon setting often returns the "real" answer of one.

[202] Notably, the adversary would (hopefully) know the epsilon values in this case, given that differential privacy is inspectable. *See* Dwork et al., *supra* note 122, at 2; *see also supra* note 79 and accompanying text.

[203] *See supra* Section II.B.1 (discussing how increased accuracy produces reduced privacy, and vice versa).

not high enough for differential privacy. If instead the real answer to this question were 5,000 then the sampling of likely outputs becomes more practical.[204] These answers appear much closer—yet are still privacy preserved—to the real answer.[205] This example, therefore, highlights the non-panacea nature of differential privacy: it is workable only in certain settings with certain assumptions, one of which is that large numbers are needed to maintain accuracy in the face of the type of noise differential privacy requires. If granular accuracy is a necessity, differential privacy may not be the best tool for the job.[206]

Finally, a likely counterargument would be that using an epsilon value of 1.098 assuming a real answer of 5,000 nonetheless appears privacy preserving, with answers like 4,998 and 4,999. That these responses are possible, however, does not affect the type of information an attacker would be learning from viewing these responses. A 1.098 epsilon value gives the attacker more assurance that any answer provided will be closer to the real answer, a feature we quantify with step one's guess difference. This high guess difference amount is validated when we look at the sampling of responses the mechanism would likely provide, averaging out to 5,000. Though an arguable position, the knowledge learned by the attacker from witnessing outputs with this particular epsilon value is likely to lead an attacker to learning too much (according to HIPAA) about how many individuals in this dataset have Chron's disease, making it more likely that HIPAA would not approve this type of data sharing.

In summary, HIPAA would likely not permit the sharing of data under a mechanism using a 1.098 epsilon value, but likely would permit the sharing if the mechanism instead used a .08 epsilon value. The risk of reidentification—guess difference—at 25% is too high for the statute to stomach, but a 2% risk of reidentification—a no-greater-than 2% boost in confidence—is likely to see a green light. In this way, this mechanism is: (*HIPAA*, .08)-differentially private.

C. Grease

Finally, it is worth being explicit about a few of the advantages, and limitations, our two-step test provides. True enough, the above analysis permitting the sharing of health data using a (.08)-differentially private mechanism is contrived. This hypothetical may not, and likely does not, fully comprehend the nuances of a legally sufficient[207] case if such a case were to arise organically. Additionally, the application of mathematical answers to legal questions, and particularly mathematical answers to statutory deidentification-type questions, has been opposed, including by the U.S. Department of Health and Human Ser-

[204] True enough, relying on the outputs to determine the appropriateness of the mechanism as compared to a statute is not rigorous, but it does help in understanding. Additionally, these numbers would need to be rounded in a post-processing step.

[205] This is the same concept as was seen in Alice's random age-generator. Section II.B *supra*.

[206] *See, e.g.*, Fredrikson et al., *supra* note 73, at 27.

[207] By legally sufficient, the suggestion is as in meeting the requirements for a case such as injury in fact, causation, and redressability. *See* F. Andrew Hessick, *Standing, Injury in Fact, and Private Rights*, 93 CORNELL L. REV. 275 (2008) (discussing standing).

vices explicitly,[208] and with good reason—ambiguity is often helpful when debating these types of questions.

That said, this data is out there; it is being purchased and sold right now. Like a spile to a tree, companies are siphoning off profit from personal, sensitive data in any and every way that is monetarily feasible. What this means for privacy is similar to what Deep Blue meant for Kasparov, the game has changed; a brave new reality has already taken hold.

The surveillance economy should not be ignored by the same statutes that seek to protect it, albeit in disjoint dollops. Instead, public policy should directly embrace the data-driven economy with the aim to promote clear, black-letter guidance. That the above test lacks a dose of justiciability does not offset the fact that it allows a data steward to reason about a privacy preserving mechanism as it would be measured against a statute—in turn giving rise to compliance-inspired confidence, something with monumental side effects for societal advancement, which are worth noting explicitly.

For one, this confidence encourages the liquidity of privacy-protective data, allowing for breakthroughs in science and technology with reduced privacy harms. Second, the test permits clear guidance on exactly how much sanitization to require for data sharing to become legal, which has additive incentives when paired with the iterative approach to common law. More specifically, when interpreting statutes, assuming no further administrative guidance is offered, the law builds on itself iteratively (e.g., the common law is marked with judicial precedent developing an understanding in a particular area). In this way, an epsilon value applied to a specific statutory situation may be deemed legal, in turn offering standard-setting effects. This is the same process that less mechanical concepts undergo: under the Fourth amendment, police can conduct a stop and frisk of someone on the street if there is reasonable suspicion of criminal activity, but police cannot conduct a stop and frisk if there is no reasonable suspicion of criminal activity. Differential privacy would run similarly: under HIPAA, you can release data using an epsilon value of .08, but you cannot release data using an epsilon value of 1.098. This would provide guidance to the idea of societal decision making in setting epsilon, as Professor Dwork has emphasized in her work on epsilon and risk-balancing decisions.[209]

[208] *See, e.g.*, *Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule*, U.S. DEP'T HEALTH & HUM. SERVS. (Nov. 6, 2015), https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html [https://perma.cc/7ZBV-9UCN] ("There is no explicit numerical level of identification risk that is deemed to universally meet the 'very small' level indicated by the method. The ability of a recipient of information to identify an individual (i.e., subject of the information) is dependent on many factors, which an expert will need to take into account while assessing the risk from a data set. This is because the risk of identification that has been determined for one particular data set in the context of a specific environment may not be appropriate for the same data set in a different environment or a different data set in the same environment. As a result, an expert will define an acceptable 'very small' risk based on the ability of an anticipated recipient to identify an individual.").

[209] *See* Dwork et al., *supra* note 122.

Differential privacy is well equipped to do exactly what it says it will do, mathematically speaking. As defined, the concept happily and routinely meets the guarantees it espouses—an *ε*-differentially private algorithm acting on a dataset with one gingersnap eaten versus the same algorithm acting on a dataset with no gingersnaps eaten produces a very similar output. Assuming you were the one who ate the gingersnap, and you knew that anyone could inspect the differentially private mechanism's source code and access its outputs, you may be assured that the chance you will be reidentified is low—an e^ε type of low. What this means for a cookie-eating regulatory statute, however, is anything but well defined.[210]

This Article introduces a novel, two-step test which may be easily applied to statutes regulating data. Step one looks at the best-case scenario for an attacker, that is, someone who, ultimately, wants to reidentify the gingersnap-eating epicure. The result of this first step is a single percentage, a legally comparable quantity representing the risk of reidentification an *ε*-differentially private mechanism accommodates. This percentage may then be measured against step two—the highest risk of reidentification a statute permits. If step one is lower than or equal to step two, the mechanism may be deemed (*statute*, *ε*)-differentially private. For example, if a court were to deem HIPAA as permitting a no more than a 2% reidentification risk (i.e., setting epsilon at .08, for a guess difference of 2%) then a mechanism could be deemed compliant: (*HIPAA*, .08)-differentially private.

That this outcome may not perfectly capture the exact percentage the legislature had in mind when it used the "very small" language found in HIPAA is beside the point; to be sure, a court, possibly with the help of an expert witness, will be able to assess the risks and weigh the benefits of data release given a justiciable case. Rather, the true benefit for this type of test lies in its ability to provide a black-letter line to data stewards, a line which is able to tout the same guarantees differential privacy touts—giving rise to confident, safe, and useful data sharing. The law and legislature, in turn, would be well served to grease the wheels on this type of compliance-inspired confidence, not continue to hinder technological progress by shielding data behind ambiguous requirements without a definitive means of meeting those requirements.

[210] Also undefined by differential privacy, though researchers are beginning to make headway, is what this means to the gingersnap-eating epicure, and whether this quantification of privacy accords with the epicurean's own standards for privacy. *See* Cummings et al., *supra* note 123; *see also* Smart et al., *supra* note 122.