Between Privacy and Utility: On Differential Privacy in Theory and Practice

Jeremy Seeman and Daniel Susser

Penn State University
Contact: jhs5496@psu.edu

Abstract:

Differential privacy (DP) aims to confer data processing systems with inherent privacy guarantees, offering stronger protections for personal data. However, thinking about privacy through the lens of DP carries with it certain assumptions, which—if left unexamined—could function to shield data collectors from liability and criticism, rather than substantively protect data subjects from privacy harms. This paper investigates these assumptions and discusses their implications for governing DP systems. In Parts 1 and 2, we introduce DP as a mathematical framework and a sociotechnical system, using a hypothetical case study to illustrate substantive differences between the two. In Parts 3 and 4, we discuss the way DP frames privacy loss, data processing interventions, and data subject participation in ways that could exacerbate existing problems in privacy regulation. In part 5, we conclude with a discussion on DP's potential interactions with the endogeneity of privacy law, and we propose principles for best governing DP systems. In making such assumptions and their consequences explicit, we hope to help DP succeed at realizing its promise for better substantive privacy protections.

1 Introduction

Formal privacy frameworks, such as differential privacy (DP)[1] are increasingly prominent, both as an approach to privacy engineering[2] and in discussions about privacy law.[3] In the academic literature, DP is usually depicted as a sophisticated new tool—an upgrade to anonymization techniques of yesteryear, which proved incapable of protecting personal data against database reconstruction attacks.[4] By injecting statistical noise into datasets, DP renders them robust to such attacks, making it difficult to infer information about the

[1] Cynthia Dwork et al., "Calibrating Noise to Sensitivity in Private Data Analysis," in *Theory of Cryptography Conference* (Springer, 2006), 265–84.

[2] Cynthia Dwork, Aaron Roth, and others, "The Algorithmic Foundations of Differential Privacy.," *Found. Trends Theor. Comput. Sci.* 9, no. 3–4 (2014): 211–407.

[3] Micah Altman et al., "What a Hybrid Legal-Technical Analysis Teaches Us about Privacy Regulation: The Case of Singling Out," *BUJ Sci. & Tech. L.* 27 (2021): 1.

[4] Irit Dinur and Kobbi Nissim, "Revealing Information While Preserving Privacy," in *Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, 2003, 202–10.

individuals they describe.[5] Thus, as advertised, DP claims to enable more useful data analysis with less risk, making hard choices about how to balance privacy and utility supposedly easier to navigate.

But technology studies scholars have long reminded us that every instrument frames the problem it sets out to solve in a particular way.[6] This is especially true for cryptographic tools, which explicitly configure power relations by allocating trust and access to different parties.[7] In this paper, we explore how DP frames these questions and who benefits from that framing. In doing so, we don't mean to minimize the value of tools that enable data sharing while protecting against database reconstruction. We argue, however, that thinking about privacy through the lens of DP carries with it certain assumptions, which—if left unexamined—could function to shield data collectors from liability and criticism, rather than substantively protect data subjects from privacy harms. As engineers use DP to architect and build data processing systems, and as policymakers turn to DP as a way of conceptualizing the normative demands of privacy and data protection, we ought to be attentive to the shifts in power and responsibility that comes with it.

The rest of Part I offers a brief overview of DP—both as a mathematical privacy framework and its implementation in sociotechnical systems—and Part II sketches a hypothetical case study illustrating how DP's framing effects create gaps between these two perspectives.

In Part III, we show how DP's framing and abstraction choices privilege certain forms of risk management that can be favorable to data collectors. First, we show that by focusing attention on data release mechanisms, DP orients privacy analysis in a predominantly forward-looking direction, obscuring potential harms enacted in data collection and database construction. In doing so, DP insulates data collectors from criticism and creates barriers to effective auditing processes necessary for holding data collectors accountable. Second, we argue that the act of setting "privacy loss budgets" (PLBs, or DP parameters that quantify disclosure risks) flattens social contexts[8] and renders the benefits of data analysis more salient while obfuscating the risks. Finally, we argue that the risk accounted for by PLBs centers individual privacy harms at the expense of collective privacy harms.

Part IV discusses how DP exacerbates well-known problems with informed consent and privacy self-management.[9] In DP frameworks, both the nature and scope of disclosure risks depend on specific properties of the relevant database, but DP only allows for transparency regarding the release mechanism and PLB parameters—not information about the database itself. Moreover, even if (theoretically) this gap was closed, empirical research demonstrates that the average person cannot easily or naturally express their privacy

[5] Cynthia Dwork et al., "Exposed! A Survey of Attacks on Private Data," *Annual Review of Statistics and Its Application* 4 (2017): 61–84.
[6] Madeleine Akrich, "The De-Scription of Technical Objects" (MIT press, 1992).
[7] Phillip Rogaway, "The Moral Character of Cryptographic Work.," 2015.
[8] Helen Nissenbaum, *Privacy in Context* (Stanford University Press, 2009).
[9] Daniel J Solove, "Introduction: Privacy Self-Management and the Consent Dilemma," *Harv. L. Rev.* 126 (2012): 1880.

preferences in the language of DP.[10] It thus provides cover for data collectors to present lopsided evidence for DP's efficacy in protecting against privacy harms.

Finally, Part V argues that effective governance for DP systems requires grappling with how DP frames these relationships between data subjects, data curators, and data users. Left unaddressed, DP could become another occasion for data curators to engage in a kind of "privacy theater," performing compliance with privacy regulations instead of fulfilling their more substantive goals, mirroring systemic problems with privacy regulation as a whole[11]. Thus, to conclude, we propose DP governance principles aimed at addressing the framing effects of DP, highlighting possible interventions both within the confines of DP and beyond the confines of DP.

Two notes before diving in. First, some of the issues we raise in what follows are common to other statistical disclosure limitation (SDL) techniques, of which DP is only one variant. We focus on DP, in particular, for several reasons: (1) in order to illustrate, with some specificity, issues that arise when mathematical privacy methods are applied in the real world, it is necessary to home in on a particular suite of formal approaches, rather than rely on generalities; (2) because DP is technically complex, its theories and methods are rapidly evolving, and it is utilized by a variety of actors (some more and others less equipped to reason carefully about its uses and limitations), it is likely to be a significant source of confusion; and (3) DP's widespread adoption and increasing prominence in academic, industry, and policymaking circles make it especially important for people to engage with critically. The stakes of getting DP right are high.

Second, the normative assumptions and social/political effects we describe below do not follow *necessarily* from the application of DP in practice. They flow, we argue, from the way DP frames privacy problems and solutions. Which is to say, unless these assumptions are made explicit, so data users and data subjects can contemplate and address them, DP will likely incline people to understand and approach privacy in certain (sometimes unhelpful) ways. Our aim in articulating these implicit framing devices and their normative assumptions and implications is—again—not to call into question DP's value and contributions to strengthening privacy overall, but rather to render concrete the real-world consequences of DP's mathematical abstractions, and in doing so, to help DP proponents deploy it carefully and effectively.

1.1 DP as a Mathematical Formalism vs DP as a Sociotechnical System

Before going further, it is important to understand that the term DP has come to describe both a set of mathematical techniques (what we'll call "DP math") as well as the larger sociotechnical systems in which they are embedded (what we'll call "DP systems"). As a mathematical framework, DP defines privacy as a set of mathematical properties of

[10] Rachel Cummings, Gabriel Kaptchuk, and Elissa M Redmiles, "'I Need a Better Description': An Investigation Into User Expectations For Differential Privacy," in *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, 2021, 3037–52.
[11] Ari Ezra Waldman, "Privacy Law's False Promise," *Wash. UL Rev.* 97 (2019): 773.

database queries. Which is to say, it offers tools for generating aggregate statistics about information contained in a database, without leaking information about the individuals whose data comprises it. A release mechanism (i.e., any function designed to answer a particular query) will satisfy DP's mathematical requirements if the result is robust to changes in one individual's record. In other words, if two databases differ with respect to one individual's record, then the query responses under DP math are nevertheless similar with high probability—DP obscures the contribution of any single person's data to the overall dataset. That obscurity is the "privacy" in "differential privacy."

For any query, this similarity is typically quantified by a numeric parameter called the "privacy loss." Smaller privacy losses ensure outputs that are closer together—more difficult to distinguish—creating stronger privacy guarantees. To satisfy DP math for a particular privacy loss, DP injects randomized noise into the results of database queries, with smaller privacy losses (i.e., more privacy) requiring more noise. Importantly, each query (and subsequent privacy loss) degrades the system's overall privacy guarantees. To keep track of these losses over a series of queries, DP composes them into a cumulative, global privacy loss budget (henceforth "PLB"), or total allowable losses under a collection of multiple queries. The PLB dictates the privacy guarantees of multiple queries at once, each query consuming some finite amount of privacy loss.

Some important theoretical findings in statistics motivated this approach. Before DP was introduced, many methods for designing mathematical data privacy protections took individual databases as their unit of analysis.[12] As a result, privacy-enhancing "data sanitization" methods were directly tailored to each specific database of interest (e.g., removing personally identifying information or pseudo-identifying information). Despite their utility, these methods proved vulnerable to database reconstruction attacks—attempts to infer individual records contained in a database. DP was developed as a response to this threat. As access to the data and computing power needed to execute reconstruction attacks has grown, such attacks have become more realistically achievable, rendering even summary statistics about databases a threat to individual privacy. For example, although the U.S. Census does not release any record-level data, Bureau statisticians were able to reconstruct individual level records from published tabular summaries, and they were able to link those records to commercially purchased data.[13]

In response to these challenges, DP (in its original formulation) introduced two important conceptual shifts. First, DP *requires* adding noise to every released result. This differs from previous disclosure control methodologies, which sometimes used privacy-preserving randomized noise but did not require it. Second, because any statistic (noisy or not) can enable reconstruction, DP focuses on minimizing *relative* disclosure risks rather than *absolute* risks. Which is to say, DP quantifies how disclosure risks change *relative to what an adversary might know before seeing the query answer*. As we discuss in what follows, DP

[12] Anco Hundepool et al., *Statistical Disclosure Control*, vol. 2 (Wiley New York, 2012).

[13] Simson Garfinkel, John M Abowd, and Christian Martindale, "Understanding Database Reconstruction Attacks on Public Data," *Communications of the ACM* 62, no. 3 (2019): 46–53.

is thus best understood as a "harm reduction" project, rather than as a tool for definitively anonymizing/de-identifying datasets—it makes disclosure less likely, not impossible.

These innovations give DP desirable mathematical properties not shared by previous approaches to quantitative data privacy: e.g., they enable methodological transparency (disclosing the way in which noise is injected into the statistics does not itself change the PLB or any resulting privacy guarantees), release mechanisms are robust to post-processing (one can use DP-generated statistics as inputs to further statistical analysis while retaining the privacy guarantees of the original release), and DP outputs *compose* (if two different releases satisfy DP math for two different PLBs, then releasing both results simultaneously satisfies DP-math for some new, typically larger PLB).[14] For these reasons, DP techniques have been incorporated into a wide variety of data-driven tools—database systems engineered to respond to queries using one of these formalisms (henceforth "DP systems").[15]

Important for our purposes, DP math abstracts away from numerous concrete design choices required when implementing DP in real-world systems. Or, to put the same point the other way around, engineering DP systems requires answering questions that DP's mathematical formalisms ignore. First, data curators (i.e., designers and operators of DP systems) must determine who their intended data users are and what requests these users want to make of the confidential data. Second, data curators must establish which database queries they will (and will not) respond to. Third, data curators must allocate components of their PLB to these different queries. Although all three of these tasks are essential to establishing a DP system, DP math typically presumes these questions are answered a priori. Yet in reality they are rarely established in clean-cut terms, which, as we will see, creates a number of underappreciated problems.

A central aim of what follows is to highlight the tensions created when attempting to establish DP systems, exploring what happens when neat mathematical formalisms make contact with real life. One benefit of mathematical framings is that they offer closure—we can prove or disprove whether the mathematical properties of a release mechanism satisfy a particular privacy definition under a set of established assumptions. No such closure exists, however, when analyzing DP systems, since we can't definitively establish whether the assumptions embedded in the mathematical abstractions map onto any system in use. Real-world applications are unavoidably messy. Thus, there's friction between DP's theorists and its systems engineers. Our goal in this paper is not to resolve such tensions but to investigate their normative implications.[16]

[14] For example, epsilon-DP has *linear* composition, wherein if one release has a PLB of X and another release has a PLB of Y, then releasing both has a total PLB of X+Y.

[15] Indeed "differential privacy" can refer to the original mathematical definition (what we have called "DP math"), any one of hundreds of other mathematical definitions of privacy with similar semantics and abstractions. See Damien Desfontaines and Balázs Pejó, "Sok: Differential Privacies," *ArXiv Preprint ArXiv:1906.01337*, 2019.

[16] In this paper, we treat DP-systems very broadly, as any privacy-preserving data processing system designed to deliver DP-math's desirable properties. Namely: relative privacy guarantees, the necessity of

Finally, to stave off potential confusion, a brief comment about terminology: privacy is a highly—perhaps "essentially"—contested term Debates over its precise meaning and value have raged for decades and continue in contemporary technical, philosophical, legal, and policy arenas.[17] In this paper, we mostly use the term "privacy" in the way technical scholars studying DP and other formal privacy methods do: as a property of data processing outputs that enables protection against unwanted disclosure of personal information.[18] We realize that this is a narrow understanding of a rich and multifaceted concept. A central goal of what follows is surfacing how, in order to make privacy amenable to mathematical precision, DP abstracts away from its social and normative complexity.[19] Still, viewing DP math as a uniform improvement over prior techniques could tempt us to think these normative issues are resolved, when in fact they are merely bracketed in a way that hasn't attracted sufficient scrutiny.

2 "Minding the gap" Between DP math and DP systems

To set the table, imagine the following:

> *A large hospital system is studying an emerging genetic disease in a small population of patients (henceforth "data subjects"). Their researchers (henceforth "data users") partner with a healthcare data technology company (henceforth the "curator") to analyze the patients' data using DP. After a few rounds of back-and-forth negotiation on query selection and privacy loss budgets, the DP-system parameters are established. Then, the patients are told that DP will allow the researchers to perform their DP-protected data analyses without revealing personal information about their disease*

randomized noise, privacy quantification through PLBs, and robustness to database reconstruction and post-processing. There are hundreds of mathematical definitions of DP that relax and strengthen the assumptions necessary for quantifying privacy guarantees in different ways (some of which might address some of the issues we raise in what follows). We focus on these properties because they align with the most frequently used definitions in publicly facing implementations of DP-systems (such as epsilon-DP, epsilon-delta-DP[16], rho-ZCDP[16], etc.). Since our interest is in the sociotechnical effects of DP systems, we focus less on the etymology or axiomatic interpretations of any one privacy definition or formalism, and more on developing a bird's-eye view of the properties most common across the most publicly visible DP systems. See Damien Desfontaines, "A List of Real-World Uses of Differential Privacy," *Ted Is Writing Things* (blog), 2021, https://desfontain.es/privacy/real-world-differential-privacy.html. We acknowledge that DP—as a mathematical approach to privacy—is but one perspective in a broad landscape of theoretical formal privacy (FP) methods. Although we are focusing on common practical use cases, we explicitly avoid a technically motivated mathematical essentialization of what DP is, distancing our work from any conception that we seek methodological closure on setting the boundaries of differential privacy.

[17] Mulligan et al., "Privacy is an Essentially Contested Concept"

[18] In the technical literature, this is sometimes referred to as "privacy of the output," as opposed to "privacy of the computation," which encompasses cryptographic and multiparty computing techniques for secure data sharing and processing. See Yehuda Lindell and Benny Pinkas, "Secure Multiparty Computation for Privacy-Preserving Data Mining," *Journal of Privacy and Confidentiality* 1, no. 1 (2009).

[19] Note that such bracketing occurred before the advent of DP, too, as both traditional privacy frameworks and DP-math work in service of the same goal, quantifying disclosure risks. See, e.g., Aleksandra Slavkovic and Jeremy Seeman, "Statistical Data Privacy: A Song of Privacy and Utility" (arXiv, 2022), https://doi.org/10.48550/ARXIV.2205.03336.

> *status, describing to them what data processing operations will take place. Once the patients consent, the researchers analyze and publish their findings on the rare disease.*
>
> *…that is, until things start going wrong. A few months after the system is introduced, two problems emerge: First, an insurance company (or "adversary"), combines the published results with consumer spending data to identify people with the rare genetic disease. Given their higher financial risks due to potentially extensive healthcare costs, the insurance company denies the data subjects coverage. The revelation creates a public relations fiasco, leading many to choose sides on who was accountable.*
>
> *The data curators argue the DP system was working as theoretically intended, and the inferences the adversary made would have been similar with or without seeing the researcher's results. Furthermore, they argue that the adversary's actions were based on predictions of someone's insurance risk, not their risk as attributable to having the rare disease or not. Therefore, it could be argued that people without the rare disease were equally affected by the decision to incorporate the data into their risk model.*
>
> *The harmed data subjects argue that, on the contrary, because they received discriminatory treatment based on their disease status, which they preferred to keep confidential, this realized harm stemmed from a substantive violation of their privacy. Additionally, the data subjects argue that knowing how the technology was used could have plausibly changed whether they consented to participating in the study or not.*
>
> *The second problem is the researchers' findings turn out to be less statistically sound than they had hoped: a nationally sponsored clinical trial fails to reproduce the DP system's results, delegitimizing the preliminary work of the hospital researchers. This second controversy creates a different set of debates in the scientific community about ensuring DP results are useful.*
>
> *Proponents of DP argue that because the release mechanism can be transparently disclosed, it is the responsibility of data users to adjust their downstream inferences to account for the errors introduced to preserve privacy. If the resulting inferences are too weak to be useful, the solution is to increase the PLB (i.e., make it easier to learn about individual patients) or design a more optimal release mechanism. The privacy loss is justified, they argue, by the more robust science it enables.*
>
> *But DP detractors argue that scientists are not trained to perform this kind of normative calculus, weighing ethical privacy concerns against the need for statistical reliability. For the researchers, it's a scandal that the PLB was not set to allow for appropriate statistical power in the first place. They argue that having only released parameter estimates from a few models, it would be computationally implausible (though possible) to use that information to reveal anything sensitive about individual patients.*

In this hypothetical scenario, the choice to use DP was motivated by two goals: (1) limiting disclosure risks for research participants, while (2) preserving enough of the data's information content to allow researchers to produce useful (i.e., statistically valid, reproducible) inferences. Despite these good intentions, neither goal was met. However, this itself isn't notable; any SDL technique could just as easily fail to meet these goals. What *is* notable is how DP framed arguments for contesting and negotiating social values that past techniques did not. In the remainder of this paper, we unpack what went wrong, paying special attention to the way DP frames considerations about privacy, utility, and other values.

First, we argue that DP math centers PLBs as the primary site where privacy is negotiated. Through this lens, it can appear as though both the privacy guarantees for data subjects and the statistical utility of the researchers' results flow entirely from this choice. But PLBs are only one element of a larger, more complicated story. Many decisions leading up to the setting of a PLB—including, importantly, the decision to use DP at all—determine how risk and utility are balanced. Specifically, in this case, the effects of choosing a particular privacy budget are more visible to the researchers than the patients ("the allocation problem", section 3.1), and DP's abstractions can disguise the curator's role in creating disclosure risks ("the responsibility problem", section 3.2). Furthermore, this helped entrench an emphasis on individual harms when the harm rendered for the patients was collective ("network problem," section 5.1).

Second, DP creates unique obstacles to "privacy self-management," the regulatory approach that makes each individual responsible for evaluating and negotiating the terms of data collection about them. Despite sustained and well-known criticisms, privacy self-management is still the dominant paradigm in privacy law and policy, at least in the US, requiring data collectors to inform people about their data practices and then leaving it up to individuals whether to consent to them. Where DP systems are used, data subjects are forced to reason about the risks and benefits of data collection in DP's terms. Because these terms are often unintuitive, expressing subjective privacy preferences in PLBs can be difficult ("the mathematical language barrier", section 4.2). For technically sophisticated data subjects, knowledgeable enough to express their preferences in the language of PLBs, making an informed decision about whether to disclose information requires curators to provide sufficient detail about the specifics of the relevant DP system, adding to the ever-expanding, impenetrable text of privacy notices few can and do read ("extending the consent dilemma", section 4.1).

Finally, we argue why these framing effects have direct implications for DP governance. While the data curator in the hypothetical relied on DP to balance privacy harms and data utility, the system's failure revealed obstacles to accountability. Not only did formal privacy guarantees provide moral cover for the data curator, but the interventions proposed for solving the data utility problem implied further privacy losses for the participants. In this way, DP helped to create the conditions for the data processing to *appear* private without addressing substantive privacy harms ("endogeneity problem," section 5.1). To resolve such tensions, either the data curator would need to improve data subject representation within DP negotiations ("Governance within DP," section 5.2) or acknowledge that DP may not target all of the privacy risks to patients ("Governance beyond DP," section 5.3).

To be clear: DP is a major advance, addressing many shortcomings of other disclosure control methods. Our aim is not to minimize these technical achievements. Rather, we want to surface the frames, assumptions, and values DP embeds in sociotechnical systems when its methods are implemented in practice—the units of analysis, measures of risk, and terms of negotiation DP-systems rely upon and enact.

3 Framing Effects of DP Decision-making

3.1 The Responsibility Problem: On Forward-Looking Harms

The first thing to notice about the way DP frames privacy decisions is that it directs attention in one direction: toward the future. In its most common form, DP assumes that data has already been collected and stored, or the data curators have prespecified exactly which data will be collected and stored; then, the decision has been made to further release it in some form, which may be publicly available or not. The only question, then, is how to do so without exposing data subjects to too much risk of being harmed by those disclosures. That is not a bad question; and, as we've suggested, the tools DP offers for addressing it represent a true advance in privacy-enhancing technology. But we should also note the questions this future-orientation draws our attention away from: e.g., Should this information have been collected in the first place? Was it collected in the right way? Should it be stored in this database, subject to these parties' control? Who decides who gets access to the data and on what terms? Whose interests does the existence of this database serve? Focusing entirely on forward-looking harms can deflect attention from past and present harms—such as illegitimate data collection, concentrating data in unacceptable ways, and so on—and it can displace responsibility for those harms from data collectors and curators onto data subjects and users. We refer to this as the "responsibility problem."

By invoking "responsibility," our goal is to show how future-orientation fails to capture the conditions which make privacy threats realizable in the present, particularly when data curators themselves help to create those conditions. Crucially, this future-orientation is not an accidental or contingent feature of DP—it is a basic assumption encoded in DP's math. By quantifying the effects of privacy harms relative to a "state of the world" before such a release exists, DP orients privacy analysis in a purely forward-looking direction. This can create conflicts or misalignments of interests: when, for example, curators and data users operate within the same organization, curators can make design choices favorable to data users—and potentially unfavorable to data subjects—based on confidential data, only to then release the sanitized statistics to data users after ensuring they meet their needs. Moreover, such conflicts can be obfuscated from public view by way of trade secret protections or other intellectual property claims; some legal scholars have even argued public policy concerns deserve exemption from trade secret statutes.[20] In total, this results in an forward-looking view where only the harms associated with future data uses (and future data users) are concerned.

[20] Peter S Menell, "Tailoring a Public Policy Exception to Trade Secret Protection," *Calif. L. Rev.* 105 (2017): 1.

Additionally, DP math's focus on relative guarantees creates further responsibility problems through not just *when* harms are quantified, but *how*. At its core, DP is a harm-reduction project: since any statistic can lead to potential disclosures by database reconstruction theorems, we can only prevent excess harms from new statistics released into the world. This technically motivated concession enables DP to quantify disclosure risks for "arbitrarily" powerful adversaries, i.e. any possible starting point for knowledge about individuals in a database. However, this theoretical property requires that no information about said individuals is used in implementing the system, such as for choosing queries or setting PLBs. Most, if not all, organizations have previously released results about the users who've contributed to their databases without DP-systems protections. For example, if the hospital researchers had published papers describing the kinds of patients more likely to have the rare genetic disease, the insurance company could use that information in attempting to detect which insurees posed higher financial risk. In a strict interpretation, we may claim that DP-math is now practically infeasible, as we've violated the requirement of randomized noise for each query response. But such a paradox is not easily resolved, because unless an organization's data has been hermetically sealed since its existence, DP-math has never been strictly adhered to in its entirety. Therefore, curators could claim absolution for downstream effects from any statistics they released which informed potential adversaries. By considering harms in a forward-looking direction, past actions are obfuscated and thereby insulated from criticism or protest.

3.2 The Allocation Dilemma: On Setting PLBs

Next, we discuss how DP frames implementation decisions for data curators and users in a way that privileges certain means for setting privacy loss budgets. As shown in the hypothetical, setting the PLB requires significant work in translating mathematical formalisms into practical privacy commitments. Because of the way DP-math frames privacy loss through PLBs, disclosure risks are more abstract and harder to interpret. By contrast, the effect of PLB settings on data utility is more salient and easily perceived. This implicitly privileges data utility as the driving force behind how PLBs are set and allocated. We refer to this problem as "the allocation dilemma."

Privacy loss budgets are one of DP's most celebrated innovations: they allow for a level of mathematical precision when balancing privacy against other values that normally eludes policy discussions. At the same time, these neat formalisms don't convey everything one needs to know when interacting with DP systems. A complete mathematical description of a release mechanism and its associated PLB describes how statistical noise is added to the data. But this description is not specific to the actual database in question—it applies to the entire database *schema*, of which any particular database is but one instance. Some actual databases may be more susceptible to disclosure than others, depending on the relative uniqueness of sensitive attributes among data subjects. Setting the PLB thus implicitly requires reasoning about an unknown worst-case database. If, for example, the data curators in our hypothetical wanted to take a privacy-first approach, they would need to consider the degree of homogeneity or heterogeneity within the patient group. Centering the PLB risks ignoring it.

Furthermore, the relationship between the PLB's *relative* disclosure risks and measures of *absolute* disclosure risks depends on numerous database properties abstracted away from DP-math. PLB-based guarantees are typically agnostic to database size and schema complexity, but knowing these might change how we would choose to set the PLB. For example, we might think differently about the privacy risks associated with learning how many people in a room had the rare genetic disease considered in our hypothetical, versus learning how many people in the country had it. Practically speaking, there's much more we could learn about individuals in the former case versus the latter case. PLB analysis, however, treats these queries the same way, despite their being substantially different interpretations of the realized privacy risks.

Similar problems emerge when considering which record attributes are the most disclosive, and who in the database is most at risk. Data subjects have different conceptions of "worst-case" risks, bound up in their personal, subjective experiences of boundary management.[21] but different individuals in the database also bear different risks a priori, simply based on their database contributions. In our hypothetical, suppose that younger patients are more likely to have the disease than older patients. In that case, older patients in the database have different disclosure risks compared to younger patients, for whom there are more similar records. Although the PLB accounts for the worst-case scenario among these possible databases, it flattens the privacy decision space, both qualitatively and quantitatively.[22]

While focusing on PLBs obscures some of these more nuanced privacy worries, it makes the upshots of data collection and analysis—data's utility—more salient. As we've seen, DP allows for transparent disclosure of the release mechanism, meaning the exact process used to sanitize any statistic is public knowledge. This is what allows data users to account for the privacy-preserving statistical noise DP introduces when they make inferences from query results. Because the magnitude of privacy-preserving errors is a direct consequence of setting the PLB, a curator can observe an exact change in data utility associated with different PLBs. At face value, this is good: previous privacy-preserving data sanitization methods did not have this transparency property, and it enables more

[21] Daniel Susser, "Information Privacy and Social Self-Authorship," *Techné: Research in Philosophy and Technology* 20, no. 3 (2016): 216–39.

[22] We note that some technical work has addressed possible design choices that allow for finer granularity in attribute-specific risk and participatory budget design; however, such tools are not representative of the dominant use practice for PLBs, which are accumulated over collections of queries. Moreover, this user specificity also fails to account for networked risk across queries, further obfuscating these effects. Regardless of the nuances in precisely how these issues are left unresolved, PLBs capture a highly abstract form of privacy risk that flattens the essential contextual nature of information flows needed to evaluate DP-systems in practice. For one such example of technical work in this area, see Wanrong Zhang, Olga Ohrimenko, and Rachel Cummings, "Attribute Privacy: Framework and Mechanisms" (arXiv, 2020), https://doi.org/10.48550/ARXIV.2009.04013.

accurate, reproducible inferences. But the salience of utility relative to risk could also have unintended consequences.

We can view database-specific privacy and utility analyses as tools for setting upper and lower bounds on PLBs, respectively. DP's framing makes setting lower bounds easier than setting upper bounds, privileging negotiations about DP-systems led by utility concerns. For example, in many high-profile DP-systems use cases, PLBs are set using "fitness-for-use" modeling in which a set of key data utility goals are defined in advance, and the end goal is finding the smallest PLB that satisfies these data utility goals.[23] This is the kind of process the U.S. Census Bureau used to choose its PLBs when implementing DP, and other organizations have likely followed suit.[24] Fitness-for-use approaches pre-determine which tasks are deemed important enough to lower bound the PLB and determine how that lower bound should be set. Privacy, by definition, plays second-fiddle to data utility, and when the current data utility for new tasks is deemed insufficient, the default response is to propose spending more PLB, as illustrated in our hypothetical.

These framing effects shape privacy negotiations between data curators, data users, and data subjects. In many cases, stakeholder interests are misaligned; sometimes they are directly at odds. DP's methodological transparency can, like all transparency, be a tool for curators to establish unearned trust and garner soft institutional power. As Claire Birchall writes, transparency confers "cultural, political, and moral authenticity…an identity as much as a mechanism."[25] As we've seen, knowledge of the DP release mechanism and PLB alone omits crucial details about how queries were selected and how the PLB was chosen. When the interests of curators and data users align, but are orthogonal to the interests of data subjects (e.g., when a for-profit company collects data for internal use), DP's transparency can be used to maintain the appearance of neutrality, regardless of whether the PLB was chosen in a way that substantially protects individuals' data.

Alternatively, when the interests of curators and data users diverge, transparency can take on a different character. In the case of the U.S. Census Bureau, some data users, such as social scientists and other researchers, have demanded more granular and accurate data, while the Bureau (the curator) is legally obligated to preserve the privacy of data subjects in a way that bars them from providing it. The Bureau, in response, has argued that DP math's methodological transparency provides data users all they need to adjust their calculations to account for the errors it introduces. However, many data users, especially demographers working with fine-grained data, expressed concerns that this transparency was not pragmatically useful for assessing the effects of errors on downstream inferences.[26] In this way, the Census tried to use transparency as a means of assuaging the

[23] Yingtai Xiao et al., "Optimizing Fitness-for-Use of Differentially Private Linear Queries," *Proceedings of the VLDB Endowment* 14, no. 10 (2021).

[24] John M Abowd et al., "The 2020 Census Disclosure Avoidance System TopDown Algorithm," *ArXiv Preprint ArXiv:2204.08986*, 2022.

[25] Clare Birchall, *Radical Secrecy: The Ends of Transparency in Datafied America* (University of Minnesota Press, 2021).

[26] National Academies of Sciences, Engineering, and Medicine. *2020 Census Data Products: A Workshop.* https://www.nationalacademies.org/our-work/2020-census-data-products-a-workshop

data user community it ultimately alienated.[27] These two preceding examples illustrate how transparency interacts contextually with the alignment of goals between curators and data users, in both collaborative and competitive manners.

Of course, transparency can be good. It's useful for understanding and reproducing data analyses, and it helps DP resolve flaws in traditional statistical disclosure limitation methods, which required a degree of secrecy about their methodologies. But we should be attentive to transparency's other effects as well, primarily in the way they modulate discussions about PLB allocation and responsibility for privacy harms and data utility. By unpacking the way DP helps to obscure these dynamics, we can better appreciate the consequences of choices often made out of view.

3.3 The Network Problem: On Bracketing Networked Harms

As we've seen, DP aims to balance two competing goals: "privacy" and "utility." Much of our discussion to this point has focused on unpacking the "privacy" side of that equation, but it's worth examining the "utility" side too. Privacy, according to DP, is protection against individual disclosure—DP aims to minimize the risk that someone could make reliable inferences about any individual contributor to a dataset. At the same time, DP aims to *facilitate* data's utility, which it defines as enabling statistical inferences about *populations*. As Michael Kearns and Aaron Roth write, "At its core, differential privacy is meant to protect the secrets held in individual data records while allowing the computation of aggregate statistics."[28]

This solves a real problem: as illustrated in our hypothetical, researchers—especially in health fields, but also computational social scientists, digital humanities scholars, and others—are eager to use data analysis techniques to learn from all of the data generated about us. DP promises to unlock these insights—this "utility"—latent in aggregate, population-level data, while simultaneously protecting the individuals whose data is being mined. By conceptualizing privacy and utility in this way, though, DP places outside its scope protection against the risks of aggregate statistics.[29] We refer to this as "the network problem."

Privacy theorists have long argued that our privacy interests are deeply intertwined—one person can suffer from another person's disclosures. Different dimensions of this problem have been variously theorized through the lens of "networked privacy,"[30] "group privacy,"[31]

[27] Joseph Scariano and Izzy Youngs, "Balancing Utility Versus Privacy in the 2020 Census: Sentiments from Data Users," *Available at SSRN 4089888*, 2022.

[28] *The Ethical Algorithm*, p. 50.

[29] As a group of prominent DP scholars puts it, a population-level inference "is not a privacy compromise, it is science." https://differentialprivacy.org/inference-is-not-a-privacy-violation/

[30] Alice E Marwick and Danah Boyd, "Networked Privacy: How Teenagers Negotiate Context in Social Media," *New Media & Society* 16, no. 7 (2014): 1051–67.

[31] Linnet Taylor, Luciano Floridi, and Bar van der Sloot (Eds.), *Group Privacy: New Challenges of Data Technologies* (Springer, 2017).

"relational privacy,"[32] and "privacy dependencies."[33] As Solon Barocas and Karen Levy argue, one person can become implicated by another person's data—i.e., inferences about the first can be made based upon data about the second—for a variety of reasons: if they have social ties, if they are alike in salient respects, or if they are different in ways that make one stand out in relief.[34] For example, American Express famously reduced a card member's line of credit because his shopping patterns mirrored those of people who failed to pay their bills on time.[35] Thus, while aggregate statistics protected by DP might make individuals harder to identify, they do not—as Barocas and Helen Nissenbaum put it—make them any less difficult to "reach."[36]

DP's architects are sensitive to this problem. Cynthia Dwork, for instance, makes plain that "the things that statistical databases are designed to teach can, sometimes indirectly, cause damage to an individual, even if this individual is not in the database."[37] In response, DP proponents argue that such harms are simply not what DP is designed to protect against.[38] Fair enough. We must recognize, however, that these kinds of harms—harms individuals incur not from disclosing information about themselves, but rather by virtue of being similar to others who have—are a central feature of digital societies.[39] From algorithmic social sorting to manipulative targeting advertising to unfair risk assessments, DP leaves untouched precisely the harms data ethics and policy are most concerned to prevent.[40] While it's tempting to cleanly divorce privacy concerns from concerns about ethical data use, Barocas and Nissenbaum argue that privacy's contextual nature inextricably links the two.[41] Worse yet, if—as we've argued—DP's other framing effects can serve to sanction more rather than less data collection and processing, there is reason to worry that DP could

[32] Karen Levy, "Relational Big Data," *Stanford Law Review Online* 66 (2013). https://www.stanfordlawreview.org/online/privacy-and-big-data-relational-big-data/

[33] Solon Barocas and Karen Levy, "Privacy Dependencies," *Washington Law Review* 95 (2020).

[34] Ibid.

[35] "The switch was not based on anything he had done but on aggregate data." https://www.nytimes.com/2012/02/05/opinion/sunday/facebook-is-using-you.html

[36] Solon Barocas and Helen Nissenbaum, "Big Data's End Run Around Anonymity and Consent," in Lane et al. (Eds.), *Privacy, Big Data, and the Public Good: Frameworks for Engagement* (Cambridge University Press, 2014).

[37] Dwork, "A Firm Foundation for Private Data Analysis" https://www.microsoft.com/en-us/research/wp-content/uploads/2011/01/dwork_cacm.pdf

[38] As Dwork puts it, that an individual can be harmed by aggregate statistics, even if they aren't in the database, "*suggests a new privacy goal*: minimize the increased risk to an individual incurred by joining (or leaving) the database." Ibid. (emphasis added).

[39] Anton Vedder, "KDD: The Challenge to Individualism," *Ethics and Information Technology* 1 (1999).

[40] On algorithmic social sorting, see Oscar H Gandy and Oscar H Gandy Jr, *The Panoptic Sort: A Political Economy of Personal Information* (Oxford University Press, 2021). On manipulative advertising, see Daniel Susser, Beate Roessler, and Helen Nissenbaum, "Online Manipulation: Hidden Influences in a Digital World," *Georgetown Law Technology Review* 4 (2020). On the ethics of predictive risk assessments, see Daniel Susser, "Predictive Policing and the Ethics of Preemption," in Ben Jones and Eduardo Mendieta (Eds.), *The Ethics of Policing: New Perspectives on Law Enforcement* (NYU Press, 2021).

[41] "Big Data's End Run Around Anonymity and Consent"

be used, rhetorically, in a way that increases rather than decreases the prevalence and scope of these harms.[42]

To prevent that from happening, we must be just as nuanced about data's "utility" as DP is about the kind of "privacy" protection it offers. When contemplating whether to allow the production of "useful" aggregate statistics, we should ask: useful *for whom*, and useful *toward what ends*? In situations like our hypothetical, where DP is used to enable population-level inferences that inform medical research and advance scientific understanding, there is a case to be made that society at large stands to benefit. By contrast, when private firms like Google and Facebook use DP to justify collecting more data about us to facilitate more precisely targeted ads, the "utility" at issue does not have the same normative implications. Fortunately, data governance scholars are developing new frameworks for navigating these individual and collective costs and benefits, and distinguishing not just between individual and population statistics, but drawing a line between acceptable and unacceptable population-level inferences too.[43] Understanding how to productively utilize DP in that process requires clarifying its promises—both in terms of the harms it aims to protect against and the insights it aims to unlock.

4 Data Subject Participation

In the previous section, we saw how DP's mathematical abstractions can shift attention away from privacy risks and towards data utility. While these considerations are important for curators and data users, data subjects can also participate in privacy negotiations as they engage with data collectors. DP math frames their choices too, as DP models a specific set of hypotheticals: whether an individual contributes to a database or not. In this section, we unpack two key assumptions DP makes about data subjects: (1) The transparency it provides meaningfully informs the decisions data subjects make about whether to contribute data. (2) Data subjects can express their privacy preferences in the language of PLBs.

4.1 Extending the Consent Dilemma: On DP-math and Informed Consent

[42] Importantly, one could argue that this problem is endemic to the project of statistical disclosure limitation more broadly, not strictly DP. But some methods have attempted to address it by constructing measures of collective risk. As an example, consider t-closeness, a risk measure that quantifies how different the prevalence of a sensitive attribute is for a particular sub-population compared to the general population. Such a measure captures a predictive harm, in that it uses relational information about the database individuals to infer the probability of someone possessing a sensitive attribute given external information. See Ninghui Li, Tiancheng Li, and Suresh Venkatasubramanian, "T-Closeness: Privacy beyond k-Anonymity and l-Diversity," in *2007 IEEE 23rd International Conference on Data Engineering* (IEEE, 2007), 106–15. By contrast, DP does not allow such techniques, as these risk measures are functions of the confidential database released without additional randomized noise.

[43] Salomé Viljoen, "Democratic Data: A Relational Theory for Data Governance," *The Yale Law Journal* 31(2), 2021. Sandra Wachter and Brent Mittelstadt, "A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI," *Columbia Business Law Review* 2 (2019).

Data subjects often engage with DP-systems by deciding whether to contribute their data to a system which produces DP-protected releases. For example, individuals may decide whether to participate in a survey whose results are publicly published based on how their responses are published using DP. These subjects require sufficient notice about how their data will be used in the DP-system to understand their potential role as a data contributor and decide whether to consent to such an agreement.[44] However, the numerous implementation choices for DP systems obfuscate the information needed for subjects to reason about notice and consent. These problems extend known problems with privacy self-management; hence, we refer to this as "extending the consent dilemma."

To start, we highlight the positive: putting to the side concerns raised in the previous section, DP math's broader adversarial assumptions make discussing the scope of privacy risks easier. Because results under DP math are robust to post-processing, we might hope that the arduous task of anticipating harmful use cases is, to some degree, abstracted away from us. Robustness to post-processing on its own remains a hallmark feature of DP and an essential reason for its success. However, for all the reasons discussed in the previous section, simply knowing a mechanism and its associated PLB tells us little about actual risks arising in a particular context. By construction, DP math creates a gap between the information needed to understand DP's theoretical guarantees in theory and DP's guarantees in the context of a realized database.

One might ask what makes this different from other statistical disclosure limitation procedures, many of which require methodological secrecy? How could something which reveals something once secret be part of the problem? Technology advocates have long viewed transparency as a unilateral good. However, an emergent literature on critical transparency studies has demonstrated how transparency characterizes the form of organizational information flows, not the content. As a result, organizations which appear transparent can shift their institutional perceptions through designating certain messaging as "being transparent" through organizational performance.

This performance certainly mattered to the data curator in our hypothetical, who had every incentive to appear committed to the goals of both data subjects and data curators. By telling patients exactly how their data would be processed and describing the protections afforded by DP, the data curator aimed to assuage the patient's fears about their data being used against them. Similarly, by telling data users exactly which data transformations were used by DP, the data curator provided the information necessary to enable the best

[44] Even though the hypothetical counterfactual scenarios presented by DP-math map onto various real-world constructs, the mathematical properties of DP imply a particular form of data subject intervention. In statistics, causal interventions are mathematically distinct from observational scenarios, which are viewed as outcomes from a past observed event. From a purely technical perspective, the data generating scenarios as formulated by DP's original definition are best interpreted by these causal interventions and not associative outcomes, implying a connection between the mathematical and sociological forms of intervention; namely, some sociological force compels potential data subjects to contribute to a database. So even on purely technical grounds, there's evidence to view DP-math through the lens of notice, a precursor to numerous data exchanges, such as those built on consent. See Michael Carl Tschantz, Shayak Sen, and Anupam Datta, "Differential Privacy as a Causal Property," *ArXiv Preprint ArXiv:1710.05899*, 2017. .

inferences from the DP outputs. Even though neither the data subjects nor data curators got their desired outcomes, the data curator's transparency could confer reputational advantages in the conversations following the controversy—not because of what they actually did, but because of how they talked about it. Such affordances are not merely hypothetical: for example, Apple has used privacy as a key selling point in advertising its hardware and software[45]. While DP is only one of the many privacy-enhacing technologies Apple uses, technical DP researchers have criticized Apple's use of DP for large privacy budgets and insufficient practical transparency with data subjects and users[46]. This is why transparency, as a property of a message's form alone, can be a double-edged sword that can sometimes work against the goals of clearer communication within privacy contestations.

These dynamics could turn insidious if the interests of curators and data users are aligned, but diverge from the interests of data subjects—perhaps because curators and users represent the same organization, or because they are otherwise transactionally incentivized to extract personal data from subjects. In such cases, appearing transparent and supportive of privacy concerns could encourage data collection (by minimizing risks, given DP's theoretical promises), rather than reduce it.

4.2 The Mathematical Language Barrier: On Expressing Privacy Values through PLBs.

The issues of data subject rights persist beyond issues explicitly related to the informed notice and consent discussed above. Moreover, they presume a worldview in which a PLB is fixed before data subjects ever engage with a particular data collection process. In such a setting, individuals may express a particular threshold for privacy loss, and their decision to participate in a DP system depends on whether the system's PLB exceeds their personal risk tolerance. Our concern is whether data subjects are equipped to express their personal privacy preferences in this way. Empirical user research has already suggested that this task is difficult for many data users, as they struggle with understanding trust models and privacy loss budgets[47]. Here we theoretically justify this problem, which we refer to as "the mathematical language barrier."

First, individual data subjects cannot model their personal boundaries in a sociological vacuum. Privacy risks are networked—they persist relative to the uniqueness of an individual's data contribution—but PLBs only capture a relative worst-case risk, regardless of how close any one individual's contribution is to this upper bound. Data subjects with relatively unique socio-demographic characteristics are typically easier to reconstruct, as has been consistently shown mathematically by techniques in the SDL literature. In our hypothetical, patients with unique sociodemographic characteristics, like a unique age for

[45] https://www.apple.com/privacy/

[46] Tang, Jun, Aleksandra Korolova, Xiaolong Bai, Xueqiang Wang, and Xiaofeng Wang. "Privacy loss in apple's implementation of differential privacy on macos 10.12." arXiv preprint arXiv:1709.02753 (2017).

[47] See Cummings, Kaptchuk, and Redmiles, "' I Need a Better Description': An Investigation Into User Expectations For Differential Privacy"; and Aiping Xiong et al., "Using Illustrations to Communicate Differential Privacy Trust Models: An Investigation of Users' Comprehension, Perception, and Data Sharing Decision," *ArXiv Preprint ArXiv:2202.10014*, 2022.

someone with the genetic disease, may be easier to reidentify or learn about within the patient database. For individual data subjects to reason about their risks of disclosure, they must how they relate to others contained in the database, knowing such information may be inaccessible to them.

Furthermore, a data subject asked to express their tolerance for privacy loss must imagine their individual tolerance for worst-case harms, and not all data subjects are equally equipped to make educated guesses about that threshold. In our hypothetical, suppose that certain risk-averse patients opted out of the study because they were aware of how health insurers use healthcare data to price discriminate. As a result, the privacy harms in the hypothetical were distributed based partially on who had that knowledge and the ability to act on it, not necessarily whether PLBs captured individual's thresholds for privacy loss. Some may argue this is evidence of the ``privacy paradox," wherein privacy attitudes and behaviors can be misaligned; counter to such arguments, this apparent paradox turns out to conflate general attitudes with context-specific attitudes, often ignoring the systemic issues at play like the systemic availability of information that would influence privacy decisions[48]. Thus, assessing privacy risk requires the difficult work of conceptualizing networked risks and worst-case adversaries, not equally accessible to, or expressible by, all database participants. And DP makes individual data subjects responsible for it.

5 Implications for DP Governance

5.1 "The Endogeneity Problem": On the Stakes of DP governance

DP frames data privacy as a harm-reduction project: because all data processing operations create some risks, DP quantifies changes in those risks that are directly attributable to the data curator's actions. In purely technical terms, this is a consequence of the database reconstruction theorem. Yet with any governance problem based on harm, solving the underlying political issue requires determining appropriate contexts, conditions, and thresholds for comparing policy alternatives, i.e., determining when a privacy harm requires legal intervetion. In this section, we describe how DP could entrench the power of technical decisionmakers in implementing DP systems, despite the many apparent conflicts of interest we've seen so far. This bears a close connection to ``legal endogeneity" in privacy law, a term which describes how managerial approaches to complying with privacy regulations often boil down to bureaucratic box-checking over substantive compliance at the expense of data subject protections. To that end, we'll refer to this as the ``endogeneity problem."

Privacy regulation involves a multitude of actors, ranging from government agencies and lawmakers to privacy engineers. While privacy law is often formulated from the perspective of regulators, technical actors inside companies are integral to realizing its promises "on the ground"—in some cases, by using privacy-enhancing technologies like DP.[49] Should DP become a more salient tool for regulators, the role and power of technologists will likely increase, given the specialized technical knowledge required to make DP work. On its own, this is not a bad

[48] Daniel J Solove, "The Myth of the Privacy Paradox," *Geo. Wash. L. Rev.* 89 (2021): 1.

[49] Kenneth A Bamberger and Deirdre K Mulligan, *Privacy on the Ground: Driving Corporate Behavior in the United States and Europe* (MIT Press, 2015).

thing: technical experts familiar with data privacy and security could help make concrete some of the difficult, abstract trade-offs discussed throughout this paper. Our concern is about the broader structures within which such decisions are made.

For many data processors, particularly those in for-profit technology firms, compliance tasks are overseen by ethics "owners" whose efforts are shaped and constrained by the nature of their roles—mediators between the firm's "internal" goals and "external" pressure to comply with the law and social norms around privacy and related values.[50] First, ethical interventions in industry settings are typically done in a way that avoids altering core business functions. Second, when change is necessary, there is a tendency toward technological "solutionism"—the assumption that there is a technological fix to any technological problem.

Given DP's framing effects, discussed throughout this paper, it offers ethics owners an attractive technical solution to privacy problems, which require little change on the part of firms. By focusing on a narrow definition of privacy that centers disclosure risks attributable to data processing, DP brackets privacy harms associated with data collection. By making privacy harms more abstract and data utility more concrete, decisions are more likely to be utility-driven than privacy-driven. While true that any privacy-enhancing technology could be implemented in a utility-driven manner DP encourages it, since privacy loss budgets are easier to lower-bound than upper-bound. In this way, DP offers a win-win for everyone—except data subjects—by giving data curators the veneer of checkbox compliance, without seriously disrupting how organizations engage with personal information. Thus, without care and attention to how DP is governed, its impact may be less substantial than its proponents imagine.

In what remains, we outline two forms such care and attention might take. The first works *within* DP (and related formal privacy approaches), focusing on how best to implement these technologies. The second explores DP's limits.

5.2 Governance within DP

Where DP successfully captures the risks of data processing, governance ought to contend with how to make trade-offs between privacy and utility more responsive to the rights of data subjects. To that end, we propose two directions for improving DP systems, with an eye toward substantive privacy guarantees.

Concretizing the harms attributable to DP

Privacy scholars are used to hollow conceptions of privacy harms. In the words of Ann Bartow, such conceptions often suffer from "too much doctrine, and not enough dead bodies."[51] This turns out to be a systemic problem in privacy scholarship, in that many often divorce theories of privacy harms from the affective dimensions of lived privacy experiences.[52] DP's conception

[50] Jacob Metcalf, Emanuel Moss, and others, "Owning Ethics: Corporate Logics, Silicon Valley, and the Institutionalization of Ethics," *Social Research: An International Quarterly* 86, no. 2 (2019): 449–76.
[51] Ann Bartow, "A Feeling of Unease about Privacy Law," *U. Pa. L. Rev. PENNumbra* 155 (2006): 52.
[52] Luke Stark, "The Emotional Context of Information Privacy," *The Information Society* 32, no. 1 (2016): 14–27.

of privacy loss operates similarly, and in doing so makes DP a more powerful tool for shielding data curators from liability than for protecting data subjects from harm. Privacy regulation that relies on DP should find ways to recenter data subjects, clearly articulating how DP systems function in practice and making potential risks concrete and visible.

To achieve this goal, data curators need to communicate about DP to data subjects in ways that make the potential harms of data sharing as tangible as the benefits. In part, this is a language issue: they must ensure that the DP "sales pitch" is not, functionally, a mechanism for extracting consent for data collection. It's easy to colloquially describe DP as a tool that "protects data subjects from arbitrary adversaries," but this wording masks the inevitable leakage that comes from data processing, misrepresenting a harm reduction project as a harm elimination project. Other approaches involve demonstrating the kinds of protections afforded (or not afforded) by DP data processing. If data processing enables certain potentially risky inferences, they ought to be publicly disclosed. For example, when the U.S. Census Bureau wanted to demonstrate the efficacy of their privacy-enhancing technology, they executed multiple reconstruction attacks giving tangible evidence for the harms reduced while acknowledging that not all risks were eliminated. While there are many such avenues for making harms more concrete, the end goal remains to ensure privacy risks are adequately communicated to data subjects.

Balancing data subject participation with expertise

In most DP systems, consent is the only mechanism through which data subjects negotiate their relationship with the data curator. There is rarely an opportunity, e.g., for data subjects to specify the PLB that matches their individual risk tolerance. (And as we've seen, most data subjects would be ill-prepared to do so if they were given the chance.) Given the importance of query choice and PLBs in determining the real meaning of DP's protections, the concerns of data subjects should be incorporated into privacy decision-making through a collaborative, participatory governance that includes both experts and lay stakeholders. A participatory process would give data subjects some say in upper-bounding privacy loss budgets for carefully selected queries, where currently there is little forcing such constraints.

As with all participatory approaches to data governance, such an undertaking is far easier said than done. Participatory design approaches often struggle to capture nuanced values like accountability when intertwined with competing interests.[53] Similarly, skepticism towards technical expertise can degrade the legitimate value privacy engineering perspectives bring to the table.[54] These effects have played out in the Decennial Census, where both data subjects and data users fought for control over how to set DP parameters.[55] Despite these challenges, privacy and political contestation are inescapable, and we ought to embrace such contestation by ensuring data subject representation in the process.

5.3 Governance Beyond DP

[53] Christopher Frauenberger et al., "In Pursuit of Rigour and Accountability in Participatory Design," *International Journal of Human-Computer Studies* 74 (2015): 93–106.
[54] Gil Eyal, *The Crisis of Expertise* (John Wiley & Sons, 2019).
[55] danah boyd, "Differential Perspectives: How Differential Privacy Upended the Statistical Imaginaries Surrounding the US Census," 2021.

In the previous section, we outline modes of DP governance appropriate in cases where DP accurately captures relevant risks. But as we've seen, DP systems are engineered around a singular conception of privacy risk: unintentional disclosure of personal information attributable to a release process. While this focus is important, it captures only one of many possible privacy harms that can result from data processing.[56] By narrowing the scope of privacy protections to individual disclosure risks, DP systems ignore other privacy risks data governance ought to manage.

Again, as we've seen, each individual data subject's disclosure risks are bound up with—relative to—others in the database. Data governance needs to consider these relational and collective dimensions of privacy, and critical legal scholars have begun to develop frameworks for engaging substantively with these issues. For example, Salome Viljoen argues that individualized notions of data privacy "miss the point of data production in a digital economy: to put people into population-based relations with one another."[57] On its own, DP can't account for these dynamics.

There are both technical and social avenues for addressing misalignments between DP's formalisms and these broad-based privacy harms. On the technical side, new research in formal privacy methods could help address relational effects by quantifying different kinds of harms simultaneously.[58] Practically, when DP or any formal privacy technique fails to capture relevant privacy harms, regulation should ensure these techniques are not the only tools at work. Such an approach aligns with the politics of ``algorithmic realism," which distinguishes mathematically formal approaches to data processing ethics problems from substantive ethical advances[59]. Regardless of approach, understanding the limitations of DP ensures that its precise, mathematical description of privacy risk does not obscure on other, equally important privacy harms.

6 Conclusion

In this paper, we discuss normative issues in the translation of DP's theoretical properties into real-world data processing systems. To be clear, our goal is not to diminish the technical contributions of DP and associated privacy-enhancing technologies. On the contrary, we draw attention to the ways DP implicitly frames privacy protection in order to encourage more careful reasoning, discussion, and negotiation about it—to help DP succeed at realizing its promise.

[56] Daniel J Solove, "A Taxonomy of Privacy," *U. Pa. l. Rev.* 154 (2005): 477.

[57] Viljoen, "Democratic Data: A Relational Theory for Data Governance."

[58] See, for example: Zhang, Ohrimenko, and Cummings, "Attribute Privacy: Framework and Mechanisms."

[59] Ben Green and Salomé Viljoen, "Algorithmic Realism: Expanding the Boundaries of Algorithmic Thought," in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 2020, 19–31.