

# On Heuristic Models, Assumptions, and Parameters

SAMUEL JUDSON, Yale University, USA

JOAN FEIGENBAUM, Yale University, USA

Study of the interaction between computation and society often focuses on how researchers model social and physical systems in order to specify problems and propose solutions. However, the social effects of computing can depend just as much on obscure and opaque technical caveats, choices, and qualifiers. These artifacts are products of the particular algorithmic techniques and theory applied to solve a problem once it has been modeled, and their nature can imperil thorough sociotechnical scrutiny of the often discretionary decisions made to manage them. We describe three classes of objects used to encode these choices and qualifiers: *heuristic models*, *assumptions*, and *parameters*, and discuss selection of the last for differential privacy as an illustrative example. We raise six reasons these objects may be hazardous to comprehensive analysis of computing and argue they deserve deliberate consideration as researchers explain scientific work.

Additional Key Words and Phrases: computation and society; computation and law; cryptography; differential privacy; machine learning; program analysis

## 1 Introduction

The 2020 United States Census spurred debate [5, 9, 22, 24, 25, 54, 64, 68, 72, 100, 101, 112] and litigation [1, 2] on the merits of employing differential privacy (DP) [45, 46] to fulfill its legally-mandated confidentiality requirement. Some dispute focused on the distinctive method by which DP protects privacy or on procedural delays its use may have caused. But many critiques simply advanced the claim that the US Census Bureau had unreasonably accepted a significant loss of accuracy for, at best, a negligible gain in privacy. The Census Bureau and differential privacy researchers put forth compelling responses to these criticisms. Still, the need for these counterarguments evinced a curious fact. The broader debate invoked many nuanced scientific and social principles: individual privacy and the social compact; the utility vs. privacy tradeoff inherent in population statistics; federalism; bureaucratic governance and administrative law; database reconstruction and calibrated noise. Yet the defense of DP for disclosure avoidance rested in considerable part on an otherwise unassuming real number: 19.61. DP is a framework for calibrating a tradeoff between the utility of statistics and the privacy of their underlying data. At the extremes it permits both choice of maximal utility or of perfect privacy. In practice, selection of a real-valued parameter  $\epsilon$  fixes the privacy loss permitted, and with it the corresponding loss in accuracy, to somewhere in the middle. The Census Bureau set global  $\epsilon = 19.61$  [24, 25]. It was from the implications of this choice that much of the resultant controversy flowed, as the most vocal criticisms were levelled not at the idea of paying for privacy with utility, but rather at the allegedly exorbitant cost accepted by the Census Bureau.

Debates over the societal and legal governance of technology usually focus on questions of modeling: how scientists and engineers represent social or physical problems in order to formulate technical solutions. But within computing the choice of an  $\epsilon$  for DP is far from unique in how arcane technical detail can drive social consequences. These details often require significant technical expertise to understand – or to even discern the relevancy of in the first place. As another example, in cryptography it is careful modeling that motivates security guarantees; captures participants, adversarial intent, and computational capacity; and justifies the overall conclusion a construction is secure. Any such conclusion, however, may be conditioned on the use of adequate key sizes and the assumed hardness of a computational problem [57, 71]. Both of these qualifiers have been exploited to diminish the practical security of theoretically-secure systems. Export-grade cryptography

---

Authors' addresses: Samuel Judson, samuel.judson@yale.edu, Yale University, USA; Joan Feigenbaum, joan.feigenbaum@yale.edu, Yale University, USA.

was weakened through legally-mandated short key lengths [4, 37], while (almost certainly) malicious parameter generation circumvented the security arguments justifying the DUAL\_EC\_DRBG pseudorandom number generator [21, 62]. In machine learning, modeling legitimizes a data universe, a choice of hypothesis class and loss function, and the interpretation of outputs. It shapes understanding and implementation of desirable qualities such as fairness or robustness, and it gives confidence that a resulting trained model will be accurate when deployed [85, 103, 104]. However, particular definitions of fairness can have social impact dependent upon a choice of parameters intended to encode a desired equity, in a manner reminiscent of, or even directly descended from, the  $\epsilon$  of differential privacy [27, 43].

Each of these are examples of how – in search of rigorous, general results – computer scientists rely upon a proven suite of techniques for reducing choices and tradeoffs down to parameters, and for confining caveats and qualifiers to careful reliance upon certain models and assumptions. When computer scientists aim only towards traditional goals of algorithmic correctness and efficiency these methods are mostly of just scientific interest. However, for much of the research community correct and efficient is an insufficient standard, and computation must further be, *e.g.*, accountable, fair, just, explainable, interpretable, moral, legal, or politically sensible [3, 4, 12, 29, 40, 76, 79, 82, 85, 88, 90, 98, 99, 102, 103, 111]. In light of such broader norms, this otherwise innocuous suite of techniques presents a unique challenge for interdisciplinary research at the intersection of computer science with law, policy, and society, let alone for the actual practice of technological governance. As our examples demonstrate, parameters and assumptions enable computations to have consequences that arise not just from how a problem is modeled or the basic mechanism of a proposed solution, but rather from the obscure and opaque technical details of particular algorithmic techniques and their theory. For DP, the basic principle of trading some statistical accuracy to protect individual privacy can be approachable and intuitive for non-technical audiences [112]. Yet, the implications to American society and governance of the particular tradeoff encoded by  $\epsilon = 19.61$  have proven far more muddled and contentious. The ongoing groundswell of interest – both purely technical and interdisciplinary – in the interaction between computation and society motivates a critical look at what form these techniques take, and the implications they can carry for the understanding and governance of computing by non-technical practitioners.

In this work we classify three types of objects that computer scientists often use to encode consequential caveats, choices, and qualifiers into algorithms and protocols: (i) proofs that hold only over *heuristic models*; (ii) technical *assumptions* of believed rather than proven truth; and (iii) numeric *parameters* that encode complex tradeoffs through (often deceptively) simple choices. As shorthand we refer to these three classes of objects collectively as *computational alluvium*. It is an admittedly forced metaphor but hopefully an instructive one: The societal effects of computation rest on these objects, which accumulate as residue of our modeling and algorithm-design decisions. In §2 we detail these classes of objects and discuss their importance to thorough evaluation of computing. In §3 we raise six hazards that make alluvium uniquely treacherous for interdisciplinary analysis, demonstrated in §4 through a more detailed discussion of differential privacy. We then consider how computer scientists might emphasize the nature and importance of alluvium in a manner accessible to practitioners in §5, before concluding in §6.

## 2 Heuristic Models, Assumptions, Parameters, and their Implications

It is a basic principle that the societal impact of a technology bears the imprint of the scientific process which developed it, and computing is certainly no different [82, 85, 98, 103]. We are interested in artifacts which are the product of a particular pattern common in computer science research: A social or physical problem is formally modeled, and an algorithm or protocol is developed that the researcher is *just about* certain solves it. It may be that the algorithm works for a slightly different model, not equivalent to the original but “close enough” to seem justifiable. It may instead be that the algorithm seems to work,

and certainly does so if certain assumptions – which are well defined on their own terms, independent of any particular application – hold. Or, it may be that the algorithm produces an effect that could be desirable, but only if attuned to the circumstances of its use which are as yet unknown or are inherently subjective. So, the algorithm designer makes that effect adjustable and places the burden of selection onto the implementer, just as how the US Census Bureau was forced to determine an appropriate  $\epsilon$  to apply differential privacy to disclosure avoidance. In any case, the algorithm or protocol is presented as a solution to the modeled problem, and the resultant artifact – respectively, a heuristic model, assumption, or parameter – is left as a consequential detail.

**Heuristic Models.** There are two different forms heuristic models take, depending on whether our interest lies in mathematical or computational reasoning. For the former, we find ourselves unable prove some fact about an algorithm  $P$  directly. So, we instead argue with respect to some heuristic model  $\mathcal{M}(P)$ , in which (a subroutine within) the algorithm is replaced with something simpler for which we can articulate an argument. Non-standard cryptographic models, such as the random-oracle model (ROM) [13] and the Fiat-Shamir heuristic [51], are canonical examples. For computational reasoning, we have a program  $P$  we would ideally submit as input to a program  $P'$  to be computed over. No such  $P'$  is known or possible, but replacing  $P$  by  $\mathcal{M}(P)$  makes the problem tractable. Program analysis often uses this technique, for example in model checking and symbolic verification of security protocols (such as in the Dolev-Yao model) [11, 31, 32, 39]. In both contexts, the suitability of the heuristic and the confidence we derive from it relate to its fidelity, *i.e.*, the adequacy and accuracy with which  $\mathcal{M}(P)$  represents the relevant behaviors of  $P$ . However, even flawed heuristics may have value. The aforementioned ROM and Fiat-Shamir heuristic are both widely used within cryptography despite admitting security proofs for insecure constructions in certain pathological cases [26, 58]. What makes a heuristic satisfactory cannot, by its nature, ever be rigorously settled. Acceptance is a social process within the technical community, ideally buttressed with formal analysis of evidence and implications. This process may be contentious, as demonstrated by the history of the random oracle model itself [26, 56, 73].

**Assumptions.** Presumption is inherent to scientific modeling as we propose theories to explain observations; thus, all but the most abstract computer science rests on uncertain beliefs somehow. Data sufficiency in machine learning, rationality in mechanism design, or adversarial modeling in information security are examples of modeling assumptions. Their plausibility cannot be divorced from the social or physical context of the problem, their validity determines whether a technically correct computation is practically useful, and their justification determines whether it can be socially beneficial. Thoughtful consideration of modeling assumptions, whether normative or positive, is perhaps the central focus of sociotechnical analysis of computation (*e.g.*, [4, 12, 23, 44, 55, 76, 82, 84, 90, 98, 99, 103, 111]).

Our interest instead lies with a distinct class of alluvial assumptions, whose validity is instead separable from any specific practical application. These tend to be narrow statements that a specific mathematical or physical construction behaves as intended despite a lack of conclusive proof. Moreover, their validity tells us only that a computation will in practice do what the modeling intends. The distinction in form between modeling and alluvial assumptions may be subtle and further complicated by how tightly coupled they often are. Given our claim that alluvial assumptions deserve explicit attention, it is important that we be able to distinguish them from modeling assumptions. We build towards a qualitative distinguishing test through examples, which also serve to demonstrate their sociotechnical importance.

One type of alluvial assumption are well-studied yet unproven mathematical statements, believed true by researchers and used as if they are. Cryptography again provides us with canonical examples in the form of hardness assumptions [59, 86], *e.g.*, the RSA assumption [97], the Diffie-Hellman (DH) assumptions [36], and the Learning with Errors (LWE) assumption [96].

Because theoretical computer scientists do not yet have the tools to prove that there are no efficient algorithms for certain problems of interest in cryptography, many practical cryptographic constructions cannot be unconditionally proven secure [57, 69, 71]. Most designs, including, *e.g.*, all practical encryption schemes, instead rely on hardness assumptions, in the sense that there exists a rigorous proof that breaking the scheme requires invalidating the assumption. This allows study of the latter in isolation, and confidence in it justifies the claimed security.

As an example, when using the Decisional Diffie-Hellman (DDH) assumption, we assume that no practical adversary can distinguish whether a random element of some specific algebraic group  $\mathbb{G}$  is independent of two others. The assumption directly states that no efficient algorithm can behave in a meaningfully different way when the element is independent compared to when it is not.

**DEFINITION 2.1 (DECISIONAL DIFFIE-HELLMAN (DDH) ASSUMPTION).** *Let  $(\mathbb{G}, q, g) \leftarrow \mathcal{IG}(1^n)$  be an instance generator where  $n, q \in \mathbb{N}$ , and  $g$  is a generator of group  $\mathbb{G}$  of order  $q$ . For any probabilistic polynomial-time algorithm  $\mathcal{A}$  and uniformly sampled  $x, y, z \xleftarrow{\$} \mathbb{Z}_q$ ,*

$$|\Pr[\mathcal{A}(\mathbb{G}, q, g, g^x, g^y, g^z) = 1] - \Pr[\mathcal{A}(\mathbb{G}, q, g, g^x, g^y, g^{xy}) = 1]| \leq \text{negl}(n)$$

where  $\text{negl}(n)$  is eventually bounded above by the inverse of every polynomial function of the security parameter  $n$ .

The truth of this statement for a specific  $(\mathbb{G}, \mathcal{IG})$  domain does not depend on the context of its use, *e.g.*, authenticity in end-to-end encrypted messaging [83, 95]. Notably, the technical requirement that  $\mathcal{A}$  runs in "probabilistic polynomial-time" is a standard modeling assumption on adversarial capacity that underlies much of modern cryptography. This demonstrates how tightly coupled modeling and alluvial assumptions may be. While hardness assumptions are alluvial, they are only of interest because of an underlying modeling assumption about how capable our adversaries are.

The DDH assumption has been carefully studied, and there are particular domains in which we are confident in its truth [17]. But it is not proven. A stroke of inspiration could find an  $\mathcal{A}$  that violates the assumption on a domain used for a deployed cryptographic scheme and break it. Even the use of unrefuted assumptions may provide opportunities to a malicious actor. Cryptographic standardization requires the fixing of domains (and often specific instances) for schemes relying on hardness assumptions. Malicious generation can render those assumptions insufficient because of auxiliary information. Such a subversion almost certainly occurred with the DUAL\_EC\_DRBG cryptographically-secure pseudorandom number generator, which is widely believed to have contained a National Security Agency (NSA) backdoor [21, 62]. Even when we have a high degree of confidence in our assumptions, their use requires care.

As a distinctly different flavor of example, technical assumptions arise from executing programs on physical hardware that imperfectly implements a mathematical model of computation. It is inherent in all of computing, but especially acute in cryptographic engineering, formal verification, robotics, and cyberphysical systems, to assume processors, servos, and sensors appropriately transmute physical phenomena from and into mathematical information. Failures of these physical assumptions can cause great harm, *e.g.*, the 2018-19 crashes of Boeing 737-MAX aircraft that killed 346 passengers. An essential element of these disasters was the behavior of control software when readings from sensors did not accurately represent the flight dynamics [105]. At the intersection of cryptography and formal methods, side-channel attacks rely on the physically observable details of how specific software and hardware implement cryptographic schemes to steal private information by, *e.g.*, exploiting data-dependent timing and cache usage [11, 94]. Formal techniques to mitigate these attacks must assume that their protective measures will be rendered effectively when physically realized.

In each of these examples, the assumption is in some way separable from its potential uses. The DDH and other cryptographic hardness assumptions are well defined mathematical statements that would be valid objects of study even if they had no practical use in cryptography; a physical sensor on an aircraft is assumed to process information from the environment correctly, regardless of how that information is put to use for safe flight; general-purpose hardware is assumed to execute a program correctly, regardless of what the application of that program may be. We define alluvial assumptions as those that are (i) *concrete* and (ii) *self-supporting*. Whether an assumption possesses these two attributes constitutes an affirmative (though qualitative) distinguishing test.

For (i), a concrete assumption is of the form *we assume this object has property P*, as opposed to generic assumptions of the form *we assume there exists an object that has property P*. This small but important distinction is raised by Goldwasser and Kalai [59] in the context of cryptography. Generic assumptions are speculative and lack a means for constructive use. Their proposal can alter the path of research, but only concrete assumptions impact the practical use of computation. All of our prior examples are concrete. The DDH assumption for a particular  $(\mathbb{G}, \mathcal{IG})$  domain states that a specific problem is computationally hard. Physical assumptions are inherently concrete as they pertain to concrete implementations. In contrast, Goldwasser and Kalai provide examples of various generic cryptographic assumptions that we do not consider alluvial, such as the existence of one-way functions.

For (ii), a self-supporting assumption is both well defined and justifiable on its own terms, independent of how it is used. We can reason about the correctness of the DDH without any reference to cryptography, or evaluate whether a flight sensor measures accurately even if it is not connected to the rest of the avionics. In contrast the validity of a modeling assumption – the rationality of auction participants, the adversarial capacity available for attacking secure communications, the sufficiency of data to model an environment, *etc.* – is inherently tied to its deployment. It may always or never be well founded, or perhaps more likely will fall somewhere in between. But a modeling assumption can never be argued valid in general, independent of the particular context in which an algorithmic solution built upon it will be used. Simply put, modeling assumptions assert that we are solving the right problem. Alluvial assumptions imply that the solution to the problem as modeled is correct, even should that model prove useless.

An important corollary to self-support is that alluvial assumptions can transfer to entirely unrelated contexts without any necessary reconsideration of their validity. If we somehow found a use for the DDH in machine learning we could apply it with all confidence due to it from cryptography. In contrast, transferring notions of, *e.g.*, rationality or adversarial intent to ML has required evaluating whether the agents in that setting possess analogous internal motivation and reasoning as in the economic and information security contexts.

**Parameters.** Often the most conspicuously consequential alluvium, parameters allow researchers to concisely specify families of algorithms. Each family member implements the same basic design, but differs in function according to its identifying parameters. A choice of parameters then selects the family member most appropriate for a specific use case. This metatechnique allows computer scientists the flexibility to build expressive and generic theories, with nuanced application to an eccentric circumstance requiring only careful parameterization. However, it has a consequence. Parameters allow the reduction of social contention or physical uncertainty to numerical choice. And, exemplified by an algorithm for fair-ranking from the literature we will shortly describe, it can be a deceptively simple choice at that. The sociotechnical implications of parameters are therefore often more immediate than for heuristic models or assumptions. Although they are not inherent to a modeled problem, parameters are frequently the intended means for its most stubbornly subjective qualities to be fixed into mathematical terms. An example of a consequential parameter choice is that of  $\epsilon$  from differential privacy. Others appear in

the recent explosion in technical enhancements for the beneficial use of machine learning algorithms, *e.g.*, fair and robust ML.

The goal of adversarially robust machine learning is to prevent an attacker given influence over the inputs to a machine learned model from compromising its accuracy. A particular security concern is that an attacker may do so through carefully constructed perturbations indiscernable or unsuspecting to human or machine scrutiny [61, 81, 108], *e.g.*, with small stickers causing an autonomous vehicle to misread a stop sign [50]. Tsipras *et al.* [108] model adversarial robustness as training a classifier with ‘low *expected adversarial loss*’

$$\min_{\theta} \mathbb{E}_{(x,y) \sim \mathcal{D}} \left[ \max_{\delta \in \Delta} \mathcal{L}(x + \delta, y; \theta) \right]$$

for candidate model  $\theta$ , distribution  $\mathcal{D}$ , loss function  $\mathcal{L}$ , and  $\Delta = \{\delta \in \mathbb{R}^d \mid \|\delta\|_p \leq \epsilon\}$  a set of perturbations parameterized by  $\epsilon$ . Increasing  $\epsilon$  enlarges this set and captures a strictly more powerful attacker. Its choice so fixes the maximal power of an adversary the learned model is trained to resist.

We might seem to therefore want as large an  $\epsilon$  as possible. But, the effect of training against this loss is to produce a model reliant upon patterns in the data that are invariant when any perturbation  $\delta \in \Delta$  is added to an input. This stability prevents those adversarial changes from altering the classification. However, it also compromises the accuracy of the classifier on fine distinctions where similarly small changes make all the difference between two inputs of different ground truth. Enlarging  $\Delta$  through increasing  $\epsilon$  extends this tradeoff to ever larger perturbations and coarser differences. Setting  $\epsilon$  implicitly becomes a choice between prioritizing robustness or accuracy. Beyond just this intuition, the authors of [108] are able to prove the existence of this tradeoff for a pathological distribution of specific structure. They also establish it experimentally, showing an inverse relationship between  $\epsilon$  and accuracy on real datasets.

We must note Tsipras *et al.* argue that this loss in accuracy may in fact ‘result in unexpected benefits: the representations learned by robust models tend to align better with salient data characteristics and human perception.’ But both robustness and accuracy are desirable for the beneficial use of machine learning. Taken at face value, these results place them in conflict. Employing this approach to robust machine learning requires choosing through  $\epsilon$  whether to accept a loss in accuracy for some security – even when, in theory, that burden might fall disparately.

In the work of Celis *et al.* [27], fairness in ranking problems, *e.g.*, search result prioritization, is modeled by placing lower and/or upper bounds on the number of entries with a given property that can appear at the top of the ranking. Although the algorithmic approach is generic, properties are naturally motivated through diverse representation. We might, for example, require that images returned for search queries not over- or under-represent some social origin or character – the present harms of which have been searingly analyzed and critiqued [89]. The authors formalize this goal as the constrained maximization problem

$$\arg \max_{x \in R_{m,n}} \sum_{i \in [m], j \in [n]} W_{ij} x_{ij}, \text{ such that } L_{k\ell} \leq \sum_{1 \leq j \leq k} \sum_{i \in P_\ell} x_{ij} \leq U_{k\ell}, \forall \ell \in [p], k \in [n].$$

Here  $R_{m,n}$  is the set of binary matrices indicating rankings of  $m$  items into  $n$  positions,  $W_{ij}$  is the utility of placing item  $i$  at position  $j$  as determined by some arbitrary (and potentially biased) process, and for a property  $\ell \in [p]$ ,  $P_\ell$  is the set of items with that property. Most important for our consideration, the parameters  $L_{k\ell}$  and  $U_{k\ell}$  specify how many elements with property  $\ell$  must or can be placed in the first  $k$  entries. To instantiate a fair-ranking algorithm from this definition requires choice of these thresholds.

We must first consider whether it is appropriate to model fairness in a use of ranking through proportional representation with respect to a set of properties. But if so, we are still left with the choice of parameters which bound those proportions. This is a conceptually simple decision: ‘in the first  $k$  results returned, no fewer than  $L_{k\ell}$  and no more than  $U_{k\ell}$  results may have property  $p$ ’ is easily understandable for a non-technical audience. However, there is no mathematical basis on which to make this choice. It is a subjective social and political question injected into the mathematical framework through parameterization, and any beneficial effect of adopting this algorithm is contingent upon it. While the function of these specific parameters are intuitive once that social character is recognized, the more sophisticated  $\epsilon$  of differential privacy or adversarial robustness demonstrate such simplicity cannot always be expected.

Another example of parameterization, in this case from outside of machine learning, arises in DEFINITION 2.1. The concrete security of a cryptographic scheme is how much computational effort is required of an attacker for some specified probability of a successful break. It allows estimation of the time and monetary investment an adversary may expect to spend directly attacking it, and is how cryptographers ultimately interpret a ‘practical adversary.’ Concrete security is tuned through choice of security parameter for the underlying hardness assumption(s) – such as the  $n$  of DEFINITION 2.1. A major battlefield of the Crypto Wars of the 1990s was intentionally weakened export-grade cryptography. For the government of the United States, preventing the legal export of practically-secure cryptographic systems was as simple as taking a theoretically-secure design and mandating a security parameter (rendered through key length) with insufficient concrete security to prevent attacks [4, 37]. Even when we are confident in hardness assumptions and our reductions to them, concrete security is still contingent on careful parameterization.

Many computer scientists recognize that alluvium can carry significant sociotechnical implications. One of the reasons for our frequent reference to cryptography is the seriousness with which its research community scrutinizes and accepts heuristic models, assumptions, and parameters, in order to robustly guarantee security and privacy (e.g., [17, 21, 26, 56–59, 71, 73, 86, 98]). However, computer scientists possess the scientific background, mathematical maturity, accrued expertise, and, frankly, time, interest, and grant dollars to carefully consider the details of technical research. The rise in interdisciplinary collaboration and education – joint degree programs, dual-track conferences and forums, and scholarship such as [12, 76, 82, 85, 90, 103, 111] – encourages that conspicuous sociotechnical concerns, such as modeling assumptions, will see sufficient consideration. Already, e.g., the quality of data used for machine learning has drawn commentary from numerous technical and humanistic perspectives, such as in [12, 23, 55, 82, 84, 85, 89, 103]. However, broad and thorough treatment of the far obscurer alluvium is much more irregular.

Three influential articles demonstrate a lack of explicit consideration of alluvium at the intersection of computer science and law: Ohm’s *Broken Promises of Privacy* [90], Barocas and Selbst’s *Big Data’s Disparate Impact* [12], and Kroll *et al.*’s *Accountable Algorithms* [76]. All three (excellent) articles are written by technically adept and knowledgeable authors – the last including multiple active computer-science researchers. Each provides detailed, thorough, and accessible analysis of the key modeling questions and proposed technical designs within their scope. However, their treatments of alluvium – and so that the nature of various mentioned algorithms and techniques are qualified and conditional – are limited.

Ohm only reproduces a figure from [20] demonstrating that a privacy vs. utility tradeoff at the heart of his analysis can be parameterization dependent, but otherwise leaves ‘the full details... beyond the scope of this Article.’ Barocas and Selbst quote that Dwork *et al.* ‘demonstrate a quanti[t]ative trade-off between fairness and utility’ in their influential *Fairness through Awareness* [43]. No mention is made, however, of the technical conditions under which the tradeoff is provable, nor that Dwork *et al.* propose for scientists and engineers to tune it through parameterizations for constraining bias and the

learning of metrics. Meanwhile, Kroll *et al.* raise multiple areas dependent upon alluvium: program analysis, cryptographic commitments, and zero-knowledge proofs. Nonetheless, their consideration of it is limited and indirect, comprised mostly of discussion on how accountable randomness requires addressing assumptions usually taken by theorists on its physical collection.

This is not surprising. These articles focus on the law, not the technology. The audience is composed of lawyers who may not have the background for or interest in any further detail, and ample references guide those who do. The choice of how much technical detail to incorporate was, we are sure, a thoughtful, measured, and ultimately wise and good decision on the part of all the authors. But it is the clear intent of each of these articles to be a reference and resource, to help guide and shape how the scientific elements of difficult sociotechnical questions raised by privacy, accountability, and data mining are discussed and understood in legal scholarship and practice. Just because detail is withheld for good pedagogy does not mean that detail is irrelevant, and the understanding of practitioners must incorporate the implications. Alluvium often falls into a grey zone. Its exclusion from foundational work at the intersection of computer science and law is justified by its obscurity and opacity, yet it remains consequential to the ideas present in that scholarship.

In evaluating barriers to reform of social harms from computational data analysis, Barocas and Selbst write ‘[s]olutions that reduce the accuracy of decisions to minimize the disparate impact caused by [data modeling] will force analysts to make difficult and legally contestable trade-offs.’ The fair-ranking scheme of Celis *et al.* demonstrates the importance of alluvium to such choices. The base unconstrained maximization problem requires no compromise, so the ‘legally contestable’ nature of the enhanced algorithm derives entirely from ‘such that  $L_{k\ell} \leq \sum_{1 \leq j \leq k} \sum_{i \in P_\ell} x_{ij} \leq U_{k\ell}, \forall \ell \in [p], k \in [n]$ .’ Any question of its acceptability and legality would reduce to (i) the modeling decision to introduce that constraint schema, (ii) the further modeling decisions of choosing properties and classifying items to define the  $P_\ell$  sets, and (iii) the choice of parameters  $L_{k\ell}$  and  $U_{k\ell}$ . The full burden of ‘fairness’ falls upon the validity of all three. It is to the great credit of Celis *et al.* that  $L_{k\ell}$  and  $U_{k\ell}$  have such simple and approachable mathematical function. But not all fairness parameters may be so explicable, let alone all alluvium of sociotechnical importance. Establishing the full measure of such solutions as Barocas and Selbst raise requires a deeper inquiry than into just modeling or algorithm design, all the way into the mathematical esoterica.

Without care, consequential questions around alluvium may come to be viewed outside of computer science as mere implementation details, especially when it is inscrutable to non-technical observers. If so, the heavy burden and significant influence those questions can bear may be delegated to people and processes without the insight to act effectively in the public interest. Moreover, the various academic, ethical, legal, political, and societal mechanisms that govern computing will fail to capture the full challenge faced by those who manage these technologies in practice. They will be rendered at best incomplete, at worst impotent.

### 3 Six Hazards

To this point we have justified our interest in alluvium through reference to how it can bear significant risk while appearing obscure, opaque, inconspicuous, or pedantic. However, alluvium possesses other qualities which intensify the challenge to sociotechnical scrutiny it presents. We view alluvium as presenting six hazards to broad and thorough analysis – in a manner notably distinct from modeling choices and algorithm design – especially for practitioners with limited scientific expertise.

**Hazard #1: Obscurity and Opacity.** Explicitly, *while its use may have social consequences, alluvium is itself of a purely technical nature. So, its existence or basic function may not be apparent to those without non-trivial scientific understanding.*



**Hazard #2: The Pretense of Formalism.** That first hazard is only compounded by how *computational alluvium often carries a pretense of formalism* when of mathematical origin. While algorithms may be presented in mathematical terms, our common language captures their creative dimension – we *design* algorithms, we *write* programs, we *construct* protocols. They do not bring the sense of inevitability that alluvium may. Heuristic models, assumptions, and parameters often present as mathematical detail and so as inherently true or false, right or wrong, in a way the subjective or empirical nature of modeling does not. Some of these objects do have well defined truth values. Cryptographic hardness assumptions, like the DDH over a given  $(\mathbb{G}, \mathcal{IG})$  domain from §2, are well defined mathematical statements. We just have not (yet?) firmly established their truth [17, 57]. Others, especially heuristic models including, *e.g.*, the aforementioned ROM [13], are often not ‘provable’ because of theoretical technicalities. Any impression they give of formal truth is likely harmless outside scientific discourse. But many more – especially parameters – often have only a social or empirical basis. Their presentation as mathematical ideas and notation, *e.g.*, the  $\epsilon$  of adversarial robustness or differential privacy, may inadvertently launder that amathematical character.

**Hazard #3: Inaccessibility.** Technical simplicity can aid scrutiny of the sociotechnical repercussions of alluvium. The simple function of  $L_{k\ell}$  and  $U_{k\ell}$  is, *e.g.*, a great strength of the fair-ranking approach of Celis *et al.* [27] as discussed in §2. However, in general such clarity is never assured, highlighting how *computational alluvium that demand careful sociotechnical scrutiny may be technically complex, and so inaccessible to practitioners*. The  $L_{k\ell}$  and  $U_{k\ell}$  parameters avoid this hazard. Although they carry immense social depth, once that is recognized it is not hard to imagine reasonable processes – commercial, legislative, regulatory, judicial – for choosing them. They are simple enough in technical form and function to be consistent with traditional modes of resolving social contention. In contrast, for adversarial robustness the opacity of  $\epsilon \in \mathbb{R}$  makes it far harder to conceive of a selection method able to comprehensively translate its sociotechnical dimensions to and from its mathematical choice. This difference stems from the gap between the focused and transparent operation of the fairness parameters and the sweeping fullness of  $\epsilon$  as the fulcrum between robustness and accuracy.

**Hazard #4: Indeterminism and Unfalsifiability.** Each of the preceding hazards concern whether the nuances of alluvium will be adequately conveyed to practitioners. The fourth hazard works in the opposite direction – it speaks to whether computer scientists can assuage concerns raised when *computational alluvium is indeterminate or unfalsifiable*. By *indeterminate*, we mean that there is no technical notion that captures all of the failures that can result from its use. By *unfalsifiable*, we mean there is no (practical) mechanism by which its choice or acceptance may be shown incorrect on purely technical grounds. In other words, we cannot know every way indeterminate alluvium can fail, while we are unable to demonstrate that the use or choice of unfalsifiable alluvium must be wrong. Indeterminate and unfalsifiable alluvium can complicate demonstration of the (non-)existence of risk and harm from computing, which is essential for its thorough sociotechnical analysis. Although this hazard is less distinctive to alluvium, it takes on a harder edge in the shadow of its first three siblings.

Heuristic models and assumptions may be given a formal structure independent of context and so are most often determinate either formally or empirically. We can know, at least in principle, what their failures could be and the technical consequences. All heuristic models are unfalsifiable by definition. The ROM and Fiat-Shamir heuristics are as noted provably false in general [26, 58], which is why reliance upon them will always be uncertain. Assumptions are usually falsifiable. Although not able to be ‘disproven’ mathematically, most physical assumptions relevant to deployed computing may be shown invalid in relevant contexts through robust empirical investigation by engineers and natural scientists. Mathematical

assumptions are (up to the independence phenomenon) falsifiable by rigorous proof, though we leave a technical comment in an extended footnote.<sup>1</sup>

Indeterminate and unfalsifiable parameters are pervasive. For many, the subjective nature of their selection precludes any purely technical notions of falsity or failure. The  $\epsilon$  of adversarial robustness, the  $\epsilon$  of differential privacy, and the  $L_{k\ell}$  and  $U_{k\ell}$  of ranking fairness all are so. There is no inherently right or wrong choice of tradeoff between robustness vs. accuracy, privacy vs. utility, or fairness vs. (perceived) utility respectively, except under a subjective and prescriptive sociotechnical judgement. However, other parameterizations may be placed on mathematical foundations that allow empirical analysis of their validity. Tuning the concrete security of a cryptographic scheme through choice of  $n$  is an excellent example [8, 19, 28]. Some parameterizations of a social or physical character may also have determinate and falsifiable implications. For example, in the theory of machine learning, a parameterized choice of desired accuracy and confidence provably dictates the theoretically required size of the training dataset, known as the sample complexity [104].

**Hazard #5: Research Attention.** The final two hazards relate to how the broader systems built around computation manage alluvium. One concern is that *study of computational alluvium may have a limited constituency within the research community* – or, at least limited in comparison to the size of the community for the relevant theory, not all researchers in which may want to investigate such detail. An important variant of this hazard is when interdisciplinary evaluation of alluvium is required by its social or physical character and implications, but little collaboration arises. For example, and as we will discuss in further detail in §4, attention from economists, social scientists, and statisticians on parameterization of differential privacy has – despite some excellent work – been dwarfed by interest in its theoretical refinement and extension [5, 35, 46, 54, 67, 78, 110].

Computer-science researchers may believe that it is the responsibility of practitioners to do the hard applied work of developing parameterizations. Even if this particular burden shifting does not transpire, analyzing parameter choices and the reliance of constructions on heuristic models and assumptions may be seen as a secondary effort for which researchers will not be as well rewarded professionally compared to developing new theoretical results. Cryptography – which, though far from perfect, remains the gold standard on addressing these hazards – provides an example of valuing such research efforts. Hardness assumptions often receive substantial, independent attention from leading researchers [17, 96], and a small number of cryptography researchers devote a considerable portion of their research program to keeping the use of many physical assumptions and parameters honest, e.g. [15, 19, 65, 66]. This line of applied analysis has been essential for confidence in the security of deployed cryptography.

**Hazard #6: A Soft Underbelly.** Finally, *computational alluvium may form a soft underbelly in regulatory and legal mechanisms, as well as in the positive use of computation to address social problems*. Any effort to manage technology through qualitative evaluation, mandates, and procedures must be sensitive to its technical eccentricities. The examples of backdoored standards and export-grade restrictions in cryptography (see [4, 21, 37] as raised in §2) show how alluvium may be vehicles for politically motivated compromise. The converse – where the subjective or indeterminate elements of a technology are used to subvert the infrastructures we build around it – is also of grave concern.

---

<sup>1</sup>Mathematical assumptions are either true or false and so are falsifiable through reasoning *about* computation. However, computer scientists are often interested in computational proofs, where statements are formally checked *within* computation by algorithmic evaluation. There exist ‘unfalsifiable’ cryptographic assumptions, for which there is no known computational protocol for one party to convince another of invalidity [59, 86]. Applications of these – especially *knowledge of exponent* assumptions (KEA) – include the use of computational proofs within privacy-preserving designs for domains such as cryptocurrencies and program verification [14, 34, 63, 92]. Although the inability to construct such computational proofs may impact modeling, this is not ‘unfalsifiable’ in the sense that we consider relevant to sociotechnical analysis.

Without thoughtful analysis stimulated by deep technical understanding of alluvium, its careless, negligent, or malicious use might defang regulation and legislation. The assurance intended by a requirement for formal verification of a cyberphysical system depends entirely upon the validity of the physical assumptions underlying that analysis. Should a search engine operator be required to integrate algorithmic techniques to minimize bias, they might reach for Celis *et al.*'s fair-ranking algorithm, with the resultant effect fully at the mercy of choice of  $L_{k\ell}$  and  $U_{k\ell}$ . Mandates for both machine-learning accuracy and for its robustness, if the latter instantiated through an approach like that of Tsipras *et al.*, would be in direct tension, only resolvable through the contentious selection of  $\epsilon$ . Ideal parameters – to the extent we may claim they exist in some qualitative and subjective sense – may be application-specific and not amenable to global standardization as cryptographic key lengths are. Any legal finding of fact as to whether a choice of  $L_{k\ell}$  or  $\epsilon$  produced harm would require significant technical expertise. Proactive regulatory approval would further require availability of that expertise on a considerable scale, given the growing use of ML. Such problems will only proliferate alongside the ever increasing use of computation.

In general, without public infrastructure conscious of the need to evaluate and guide dependence on alluvium, it may be difficult to bring about desired social effects. Further, the nuances of alluvium are shaped by modeling, so they are only addressable after clearing a first hurdle of designing legal and regulatory systems around the intricate difficulties of the latter. The muddier that process is, the harder establishing consensus around computational alluvium may very well be. Uncertainty as to whether and how robustness should be mandated for machine learning, *e.g.*, might complicate the ultimate selection of  $\epsilon$  as stakeholders disagree on appropriate burdens and goals.

The potential of alluvium to subvert oversight also applies to any suggestion, like that of Abebe *et al.* [3], to use computation as a tool for positive social change. As they write, 'because they must be explicitly specified and precisely formalized, algorithms may help to lay bare the stakes of decision-making and may give people an opportunity to directly confront and contest the values these systems encode.' However, the uncertain and subjective nature of alluvium strikes at our ability to understand just what exactly systems built on computation encode and express. As a general principle, this final hazard states that, when the effect of a computation is dependent upon alluvium, any attempt to control and wield it for social gain is reliant upon due consideration of its heuristic models, assumptions, and parameters. This is particularly precarious in light of the first five hazards.

#### 4 Differential Privacy

We now return to a parameterization that exemplifies all of our hazards, the  $\epsilon$  of differential privacy [41, 42, 45, 46, 112]. DP provides a rigorous and provable definition of privacy for statistical database queries, as heuristic anonymization techniques are often susceptible to attacks [38, 87, 90]. The principle of DP is to associate privacy not with constraining data collection or collation, but rather data analysis.

DP is an *indistinguishability* definition derived from cryptographic theory [60]. Informally, two probability distributions are indistinguishable if no adversary can determine from which (a sequence of) observed samples are drawn. DEFINITION 2.1 provides an example. This concept extends to one of *near-indistinguishability*, where nearness is moderated through a parameter  $\epsilon$ . The justification for DP is *if the outcome of a statistical analysis is near-indistinguishable between databases that differ only in the inclusion of a single entry, then an adversary cannot learn too much about that entry from the outcome of the analysis alone*. A differentially-private *mechanism* is a statistical query guaranteed (usually through the introduction of calibrated noise) to be insensitive to any specific data entry, yet still give a reasonable approximation of the desired statistic.

The original formulation of DP makes near-indistinguishable rigorous through the following parameterized definition [45].

DEFINITION 4.1 ( $\epsilon$ -INDISTINGUISHABILITY). *Two random variables  $A$  and  $B$  are  $\epsilon$ -indistinguishable, denoted  $A \approx_\epsilon B$ , if for all measurable sets  $X \in \mathcal{F}$  of possible events:*

$$\Pr[A \in X] \leq e^\epsilon \cdot \Pr[B \in X] \text{ and } \Pr[B \in X] \leq e^\epsilon \cdot \Pr[A \in X].$$

To provide a formal theory of data privacy, DP operates over a universe  $\mathcal{D}^n$  of databases with  $n$  entries drawn from some domain, such as  $\mathbb{R}^d$ . For a given database  $D \in \mathcal{D}^n$ , an adjacent database  $D^-$  differs from  $D$  only in having one constituent entry deleted. A mechanism  $\mathcal{M}$  is then identified with a random variable taking on its output when executed over a given database.

DEFINITION 4.2 ( $\epsilon$ -DIFFERENTIAL PRIVACY). *A mechanism  $\mathcal{M}$  is  $\epsilon$ -differentially private if for all adjacent databases  $D, D^- \in \mathcal{D}^n$ ,  $\mathcal{M}(D) \approx_\epsilon \mathcal{M}(D^-)$ .*

This definition is not unique. According to a recent survey, ‘approximately 225 different notions, inspired by DP, were defined in the last 15 years’ [35]. Each variant modifies one or more aspects of DP through an alternative mathematical formulation. Many of these definitions require distinct parameterizations.

Any theory of privacy built on DEFINITION 4.1 depends absolutely on careful choice of  $\epsilon$ , demonstrated in the extreme as any two distributions on the same support are  $\epsilon$ -indistinguishable for all

$$\epsilon \geq \sup_{X \in \mathcal{F}} \left| \ln \left( \frac{\Pr[A \in X]}{\Pr[B \in X]} \right) \right|.$$

If, e.g., we let  $A$  take on a fair coin flip and  $B$  take on a biased coin flip with a 90% chance of heads – two distributions hardly indistinguishable in any intuitive sense of the word – then  $A \approx_{1.61} B$ .<sup>2</sup> Differential privacy is *prima facie* useful as it provides mechanisms that limit information leakage from statistical database queries. However, the beneficial use of DP requires principled choice of  $\epsilon$ . Only then can a deployment provide a meaningful assurance of privacy to the individuals whose data is at risk.

An oft-used descriptive name for  $\epsilon$  is the *privacy budget* [45], spent through information leakage in response to (a composed sequence of) queries. Like any budget, if too limited it has little utility – *i.e.*, the returned statistics will not be meaningful – while if too generous nothing is beyond acquiring – *i.e.*, little to no data privacy is actually provided. It is a knob that must be tuned to tradeoff privacy and accuracy, with (i) no immediate from-first-principles approach with which to do so, as evidenced by the breadth of techniques proposed [5, 46, 47, 54, 67, 74, 75, 77, 78, 80, 93], and (ii) penalties to accuracy or privacy from a flawed choice, as demonstrated by the controversy surrounding its use for disclosure avoidance in the 2020 US Census [5, 9, 22, 24, 25, 54, 64, 68, 72, 100, 101, 112], and critical comments on deployments by Uber and Apple [52, 106].

There is a simple but helpful economic understanding of  $\epsilon$ , presented in [46]. Suppose the running of a mechanism leads to some real world outcome that an individual derives some profit or loss from, such as a job offer or a higher interest rate on a loan. The guarantee of differential privacy provides that the expected utility (the average profit or loss) for the individual cannot change by more than a factor of  $e^\epsilon$  depending on whether or not their data is included in the analysis. When  $\epsilon < 1$  then  $e^\epsilon \approx 1 + \epsilon \approx 1$ , and the utility can barely change. However, when  $\epsilon$  is (much) bigger than one, suddenly this (admittedly very rough) upper bound on the worst-case change in utility can be immense. If, e.g.,  $\epsilon \approx 19.61$ , then this  $e^\epsilon$  factor is over three hundred million. Although this is just a bound, in practice it indicates that an individual may have no formal guarantee they their participation in the analysis will not drastically alter the likelihood they are harmed instead of helped by doing

<sup>2</sup>This example, while simple, is not idle. Coin flips are the basis of *randomized response*, a sociological research technique for deniable acquisition of data that is often used to motivate DP [46].

so. Invoking Tolstoy, Dwork *et al.* in [44] formulate the maxim that ‘[w]hile all small  $\epsilon$  are alike, each large  $\epsilon$  is large after its own fashion, making it difficult to reason about them.’ When large epsilons appear in practice they demand scrutiny.

The parameterization of a differentially private mechanism is not the only concern with its use, as the deployment of DP brings all the attendant difficulties of modeling. Any system that depends on statistical analysis or learning can – whether through malice, negligence, or earnestness – gild harms with a craven appeal to quantitative impartiality. Even assuming best intentions, data may be of insufficient quality or completeness for its proper use. This risk is made even more acute by the noisy nature of DP mechanisms, which require an excess of signal to survive. The use of DP for the 2020 US Census may, *e.g.*, wipe small communities off the statistical map [64]. Which of the many variants and derivatives of DP is best suited for a given setting may also be a delicate decision. Moreover, by placing the locus of privacy on data analysis rather than collection or collation, even a socially beneficial use of differential privacy opens the door to the eventual misuse of that data in the future [98].<sup>3</sup> Finally, DP is designed to limit the marginal harm to a single individual from their decision to allow analysis of their data, but makes no promises about harm to them from trends at the population level. The canonical example, presented in depth in [46], concerns differentially-private analysis relating health data and healthcare costs in order to set insurance premiums. If an individual has a behavior or condition that correlates with higher costs across the dataset, DP promises the marginal inclusion of their data into the analysis will not greatly exacerbate any increase to their premiums due to the discovery of this relationship. However, it may be that if every individual with the behavior/condition refuses to participate in the analysis, the correlation will go unnoticed and their premiums will not increase at all. Differential privacy minimizes the risk to an individual from the *additional* inclusion of their data, but does not necessarily minimize the risk to that individual from inferences made about them based on population-wide trends, to which their participation contributes. This point received particular attention during the debate over the 2020 US Census [22]

But in the end, we are left with the alluvial  $\epsilon$ . Its choice is, in the opinion of Dwork and Smith, ‘essentially a social question’ [47]. The problem has been studied formally through the lenses of resource allocation, economic valuations, and individual preferences [5, 67, 74, 75, 93]. These approaches read promisingly, but generally require a responsible party – an individual themselves, or a corporation or government acting on their behalf – to precisely quantify privacy valuations or preferences over data and events. Whether this requirement is tenable is at best contentious in the prevailing research, especially for individuals who may struggle to comprehend the basic principles of differential privacy or assess the economic value of their data [6, 7, 113]. The example of Uber also shows even technically adept corporations may struggle with scientific evaluation of DP, let alone with earning the trust to not misuse data from its subjects [52, 90]. Alternative research has explored using statistical techniques [77, 78, 80], which are however contingent on knowledge of some or all of an attacker’s goals, prior beliefs, and auxiliary information. These requirements reverse one of the most lauded attributes of DEFINITION 4.2, that the privacy guarantee of DP holds absolutely no matter the adversary or the data distribution [46].

Perhaps more realistic assessments of how  $\epsilon$  will be chosen come from Dwork and Roth [46], from the experience of the US Census Bureau [5, 24, 25, 54], and from Dwork *et al.* in *Differential Privacy in Practice: Expose Your Epsilons!* [44]. In the first, after raising both technical and procedural, *i.e.*, sociotechnical, difficulties with mathematically prescriptive techniques, the authors float an Epsilon Registry to ‘encourage better practices through transparency.’ Though such a registry might very well develop principled defaults and standards in time, the need for such a process implicitly lays bare our inability to confidently make a privacy enhancing choice of  $\epsilon$  for any given use case on its own terms. As for the US Census Bureau, the

<sup>3</sup>There do exist alternatives to the naive *curation* model of differential privacy where a trusted party handles data collection and analysis, most notably the *local* [46, 70] and *shuffle* [16, 30, 48] models. Although they do not require complete trust in a single party, they have reduced accuracy, require additional data, and/or carry increased computational cost.

words of its own researchers in [54], detailing a use of differential privacy predating its application to the census itself, are quite illuminating.

The value was set by having the practitioner prepare a set of graphs that showed the trade-off between privacy loss ( $\epsilon$ ) and accuracy. The group then picked a value of  $\epsilon$  that allowed for sufficient accuracy, then tripled it, so that the researchers would be able to make several additional releases with the same data set without having to return to [the Data Stewardship Executive Policy committee] to get additional privacy-loss budget. The value of  $\epsilon$  that was given out was far higher than those envisioned by the creators of differential privacy.

Finally, in *Expose Your Epsilons!* the authors revisit the registry, now motivated by a case study on how  $\epsilon$  has been set by researchers and engineers in various early deployments of the technology. Although the study includes examples of survey respondents who, e.g., set their  $\epsilon$  based on detailed threat modeling, Dwork *et al.* report selection methods similar to or even less principled than that of the Census Bureau, noting even that ‘there were practitioners who admitted the choice of  $\epsilon$  was completely arbitrary without much consideration.’ The authors conclude that ‘[i]n spite of the widespread preference of utility over privacy, there [is] no general agreement on how to choose  $\epsilon$ .’

**The Six Hazards.** Selection of  $\epsilon$  traverses each of our hazards. For the first, we cannot expect someone without a mature mathematical background to read DEFINITIONS 4.1 and 4.2 and immediately understand the sociotechnical implications of  $\epsilon$ . In their *Differential Privacy: A Primer for a Non-Technical Audience*, Wood *et al.* spend significant time introducing  $\epsilon$ , explaining its characteristics and implications, and experimentally demonstrating its effect [112]. However, this discussion focuses on a choice of  $\epsilon \in (0, 1]$  lying within the stiller theoretical waters where its consequences can be neatly characterized and explained. The authors leave to a footnote important real world examples – including those of Apple and the US Census Bureau – that use large  $\epsilon$ s which fall significantly above that range of values addressed in the article body. As such, in its thoughtful pursuit of a balance between depth and accessibility the main text does not in the end illuminate the real complexity of the mathematical theory or available empirical evidence. As for our second hazard, in supporting documentation for a court filing defending its use of DP, the US Census Bureau highlights that it is ‘*mathematically grounded* [emphasis added] in a way that allows statisticians to fully understand the limits of what they can make available and what kind of privacy they can provably offer’ [2]. Although this statement is factually true, it emphasizes the formal nature of DP as the essential element, while omitting that those limits are a question of social choice. Although that mathematical grounding of DP allows an understanding, it does not itself provide one. The debate around the census itself demonstrates how the social elements of that understanding can be highly controversial.

As for the third hazard, data privacy is of immense importance in the information society [90, 112]. But, DP and its variants – among our most expressive and powerful techniques for providing it – are dependent on careful management of this enigmatic parameterization. Although the simple utility analysis given previously provides accessible guidance as to why small epsilons are generally acceptable, the opaque consequences of large epsilons makes reasoning about the sociotechnical effects of their use much more difficult [44, 46]. An epsilon registry may provide a path towards consensus around choice for deployment, but the proposal of the registry itself belies our inability to always set  $\epsilon$  for a given domain on its own terms. For our fourth hazard, choice of  $\epsilon$  is unfalsifiable and indeterminate. What constitutes a privacy violation over statistical data can not always be well defined, let alone reduced to an  $\epsilon$ . A sense of privacy is often intangible, may be communal as well as individual, and may change with time. Some might also consider a given privacy violation to be outweighed by the public good and so not a failure in a sociotechnical sense. Although we may be able to show a given choice of  $\epsilon$  does not

prevent a specific attack over a specific data distribution in a specific adversarial context, when we strip away those caveats to a broader understanding of privacy violations, our ability to falsify a choice of  $\epsilon$  melts away.

For the fifth hazard, as noted in §3, work on setting  $\epsilon$  in a principled way has lagged behind efforts to extend the theory of differential privacy. A recent survey that found hundreds of variants of DP in the literature covers the selection of  $\epsilon$  in a short paragraph [35]. For the sixth and final, the importance of data privacy has made it a central focus of legislation and regulation of technology [49, 90]. Any attempt to use or manage DP in this context requires careful consideration of  $\epsilon$  on which the actual privacy enhancement depends. Otherwise, these regulatory efforts will be greatly limited in precision, if not in influence.

We stress that these hazards are only just that. Their adverse effects are not inevitable. In both theory and practice – as has been quite convincingly argued by its advocates in the case of disclosure avoidance – differential privacy can provide accurate mechanisms buttressed by concrete privacy guarantees in a provable balance all prior technical approaches to the problem lack. Its deployments for the census and beyond have almost certainly already led to meaningful privacy gains.

## 5 Explaining Alluvium

Many of our references originate in an ongoing effort by technically knowledgeable authors to proactively illuminate the scientific origins of socially consequential aspects of computation, e.g., [4, 10, 12, 76, 82, 90, 111, 112]. However, we find such work is often structured as a primer, and tends to be limited in both the breadth and depth of its scientific presentation. This limitation is natural and reasonable. An accessible narrative is essential for a non-technical audience. However, such writing may gloss over important points so as not to plunge an easy technical account into a morass of detail. Mathematical objects as small and peculiar as computational alluvium may not be given due consideration. Proactive education of the nature of alluvium requires a companion format that goes beyond such primers to catalogue – as thoroughly as is reasonable – every detail of a technology that impacts sociotechnical scrutiny.

An encouraging invention in modern computer-science research is the Systematization of Knowledge, or SoK, which has arisen within the applied cryptography and information security communities. Each of [11, 18, 33, 35, 53, 91, 107, 109] is an example on a relevant topic. The goal of these documents is to not only survey the current state of research within the domain, but to relate and organize various designs and definitions in a way that surfaces dependency, inclusion, and variation among them. In other words, to go beyond a simple survey and characterize how ideas have developed and where they lie in relation to each other. The best SoKs are structured to allow a reader with neither deep nor broad knowledge of the field to assess the current state of affairs and jump to the work most relevant to them. The recent *SoK: Differential Privacies* by Desfontaines and Pejó is an excellent example of a SoK with this structural clarity [35]. It provides seven axes along which differential privacy definitions vary, and it charts the relationships between the many proposals along them.

A possible companion to primers for educating the nature of computational alluvium would be a Systematization of Knowledge for Practitioners, or SoKfP. Such a document would survey and organize (i) the dimensions and trends in modeling of a given social or physical problem; (ii) their relationships to various published algorithmic techniques; and (iii) the dependence upon computational alluvium of those designs. Unlike a traditional SoK, the purpose of the document would not be to give a complete accounting of scientific knowledge. Rather, the goal would be to capture as thoroughly as possible the mechanisms by which scientists have proposed computation as part of solving a social or physical problem. An essential element to this narrative is the interpretation of these techniques under the most influential empirical, humanist, and social scientific perspectives. Organized and written at an appropriate level of detail – preferably by interdisciplinary scholars or in collaboration with knowledgeable practitioners – such a document could allow readers to quickly but carefully determine a

broad picture of the technical methods by which computer scientists have proposed solving a problem. This analysis would consider the repercussions of their possible deployment as evaluated through a broad sociotechnical lens.

Consider an ideal companion SoKfP to the work of Desfontaines and Pejó on differential privacy. It might begin by accounting for how compatible the underlying principle of modeling data privacy through indistinguishably-based data analysis is with perspectives from law, policy, economics, the social sciences, and the humanities, such as [90]. Any notable variants of DP – especially those that respond to concerns of modeling or alluvium with the standard definition – could then be recounted with discussion as to when one may be preferred. Careful description of these definitions in the mode of a primer similar to [112] would then motivate  $\epsilon$  and our understanding of its principled choice. This discussion would focus on – and carefully distinguish between – our theoretical, empirical, and social understandings of DP and  $\epsilon$ . Such a document would be a valuable reference to an engineer looking for guidance on how to choose  $\epsilon$  in deployment, a policymaker attempting to integrate DP into regulatory infrastructure, a lawyer trying to demonstrate harm from a loss of privacy, or a social scientist trying to interpret data analyzed under it.

*SoK: Hate, Harassment, and the Changing Landscape of Online Abuse* by Thomas *et al.* is perhaps the closest work we are aware of in form and purpose to our conception of an SoKfP [107]. It combines research from various domains and methodologies, including a significant survey, to construct a taxonomy of hate and harassment in online spaces. However, as its focus lies less on the function of internet platform technologies and more on the social behaviors they enable, alluvium mostly falls outside its scope. In a quite different way *A Hierarchy of Limitations in Machine Learning* by Malik [82] furnishes another example of work similar in purpose to our proposed SoKfPs. It is mostly composed of qualitative methods, although written by a scientific expert who uses advanced mathematical tooling at points in the narrative. Malik is intentionally opinionated ‘making [it] a work ultimately of practice and of a practitioner rather than of analysis.’ Additionally, it covers an immense scope which naturally limits the provided technical detail. Nonetheless, a synthesis in form of it with a technical SoK provides a rough outline of the structure we view as potentially conducive to effective and broadly scoped presentation of urgent topics such as differential privacy, robust machine learning, or exceptional law enforcement access.

Despite efforts by technical researchers to proactively produce scholarship helpful to practitioners, there will also always be a reactive role for computer scientists responding to scrutiny from without, such as from journalists or during expert testimony. Every query or line of inquiry practitioners may raise cannot be predicted ahead of time. Even if scientists could, systematizing it all would be an impractical undertaking. How best to convey the nuance of alluvium in an *ad hoc* manner will likely be specific to the exact combination of computer-science discipline and the expertise of the questioner. Nonetheless, our examples and hazards highlight generic attributes of alluvium whose regularized treatment may broadly help sociotechnical analysis. A SoKfP provides the opportunity to organize and present our knowledge in a manner uniquely tailored for use by specific communities of practitioners. These generic attributes are far less structured, but are at least intuitively understandable.

Most importantly, we consider the following loose classification of alluvium. It is intended to give a simple context to how an object originates and the extent to which computer scientists understand its acceptance or selection. A heuristic model, assumption, or parameter may be:

- (i) *formal*: the object has a well defined mathematical truth that has not been conclusively established or is being used in spite of its known falsity; *e.g.*, heuristic models and cryptographic hardness assumptions [57, 58, 86];



- (ii) *formal-empirical*: a formal mathematical argument provides a basis on which to conduct empirical analysis; e.g., cryptographic key sizes (security parameters) [8, 28, 71], the effect of  $\epsilon$  on the accuracy of a differential privacy mechanism [41, 42, 45, 46, 112], sample complexity [104];
- (iii) *physical*: the object represents a physical process that can be evaluated empirically by engineers or natural scientists; e.g., assumptions on sensors in cyberphysical systems [105]; choice of  $\epsilon$  in robust ML when the modeled process is physical [61, 81, 108];
- (iv) *social-empirical*: the validity or choice of object is a social question, but one that can be placed in empirical terms by social scientists; e.g., choice of  $\epsilon$  in robust ML when the modeled process is social [61, 81, 108], choice of  $\epsilon$  in differential privacy following the work quantifying privacy in economic terms [5, 67, 74, 75, 93]; choice of fairness parameters under empirical advisement from social scientific research [27, 43, 103]; and
- (v) *social*: the validity or choice of object is a social question that can only be resolved by humanist analysis of its effects; e.g., all social-empirical examples when we do not have or believe in the completeness of social scientific methods.

The distinction between the last two cannot be entirely resolved by computer scientists alone, but how naturally social scientific concepts map onto computational counterparts is an essential question technical researchers must help analyze.

To this classification we add four additional markers. First, whether the object is indeterminate. Second, whether it is unfalsifiable. Third, whether it receives active research attention and has a robust scientific literature. And fourth, whether there exists engineering experience with its development and deployment. For many practitioners, characterizations of alluvium – even without the standardized structure of a SoKfP – along these high-level lines may be more helpful than attempts to demonstrate their function through mathematical descriptions, toy examples, or experimental results from irreconcilable domains.

## 6 Conclusion

We have discussed how heuristic models, assumptions, and parameters – our computational alluvium – contribute to the sociotechnical dimensions of computation. Our thesis is, in essence, that the small and secondary nature of these objects does not mitigate their hazard to our understanding of how programs and protocols interact with society and the world. Our proposals in §5 may, we hope, stimulate computer scientists to consider how best to address sociotechnical concerns with alluvium in a manner accessible to practitioners. Regardless, it is essential that computer scientists address how to consistently provide the full measure of designs to those engaged in their principled use within society.

## References

- [1] Complaint, Alabama v. United States Department of Commerce (2021) (No. 3:21-cv-211-RAH) <https://storage.courtlistener.com/recap/gov.uscourts.almd.75040/gov.uscourts.almd.75040>.
- [2] Defendants' Response in Opposition to Plaintiffs' Motion, Alabama v. United States Department of Commerce (2021) (No. 3:21-cv-211-RAH-ECM-KCN) [https://www.ncsl.org/Portals/1/Documents/Redistricting/Cases/4\\_13\\_21\\_ALvUSDC\\_DefResponsewithAppendices.pdf](https://www.ncsl.org/Portals/1/Documents/Redistricting/Cases/4_13_21_ALvUSDC_DefResponsewithAppendices.pdf).
- [3] Rediet Abebe, Solon Barocas, Jon Kleinberg, Karen Levy, Manish Raghavan, and David G. Robinson. 2020. Roles for Computing in Social Change. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 252–260.
- [4] Harold Abelson, Ross Anderson, Steven M. Bellovin, Josh Benaloh, Matt Blaze, Whitfield Diffie, John Gilmore, Matthew Green, Susan Landau, Peter G. Neumann, et al. 2015. Keys under Doormats: Mandating Insecurity by Requiring Government Access to all Data and Communications. *Journal of Cybersecurity* 1, 1 (2015), 69–79.
- [5] John M. Abowd and Ian M. Schmutte. 2019. An Economic Analysis of Privacy Protection and Statistical Accuracy as Social Choices. *American Economic Review* 109, 1 (2019), 171–202.
- [6] Alessandro Acquisti and Jens Grossklags. 2005. Privacy and Rationality in Individual Decision Making. *IEEE Security & Privacy (IEEE S&P '05)* 3, 1 (2005), 26–33.
- [7] Alessandro Acquisti, Curtis Taylor, and Liad Wagman. 2016. The Economics of Privacy. *Journal of Economic Literature* 54, 2 (2016), 442–92.

- [8] Gorjan Alagic, Jacob Alperin-Sheriff, Daniel Apon, David Cooper, Quynh Dang, John Kelsey, Yi-Kai Liu, Carl Miller, Dustin Moody, Rene Peralta, et al. 2020. *Status Report on the Second Round of the NIST Post-Quantum Cryptography Standardization Process*. Technical Report. National Institute of Standards and Technology (NIST).
- [9] Julia Angwin. 2021. Can Differential Privacy Save the Census? Interview with Cynthia Dwork, posted May 15, 2021 at <https://www.getrevue.co/profile/themarkup/issues/can-differential-privacy-save-the-census-602525>.
- [10] Kenneth A. Bamberger, Ran Canetti, Shafi Goldwasser, Rebecca Wexler, and Evan Joseph Zimmerman. 2021. Verification Dilemmas, Law, and the Promise of Zero-Knowledge Proofs. (2021).
- [11] Manuel Barbosa, Gilles Barthe, Karthikeyan Bhargavan, Bruno Blanchet, Cas Cremers, Kevin Liao, and Bryan Parno. 2019. SoK: Computer-Aided Cryptography. *IACR Cryptol. ePrint Arch.* 2019 (2019), 1393.
- [12] Solon Barocas and Andrew D. Selbst. 2016. Big Data’s Disparate Impact. *Calif. L. Rev.* 104 (2016), 671.
- [13] Mihir Bellare and Phillip Rogaway. 1993. Random Oracles are Practical: A Paradigm for Designing Efficient Protocols. In *Proceedings of the 1st ACM Conference on Computer and Communications Security (CCS ’93)*. 62–73.
- [14] Eli Ben-Sasson, Alessandro Chiesa, Daniel Genkin, Eran Tromer, and Madars Virza. 2013. SNARKs for C: Verifying Program Executions Succinctly and in Zero Knowledge. In *Annual International Cryptology Conference*. Springer, 90–108.
- [15] Daniel J. Bernstein, Yun-An Chang, Chen-Mou Cheng, Li-Ping Chou, Nadia Heninger, Tanja Lange, and Nicko Van Someren. 2013. Factoring RSA Keys from Certified Smart Cards: Coppersmith in the Wild. In *International Conference on the Theory and Application of Cryptology and Information Security*. Springer, 341–360.
- [16] Andrea Bittau, Úlfar Erlingsson, Petros Maniatis, Ilya Mironov, Ananth Raghunathan, David Lie, Mitch Rudominer, Ushashree Kode, Julien Tinnes, and Bernhard Seefeld. 2017. Prochlo: Strong Privacy for Analytics in the Crowd. In *Proceedings of the 26th Symposium on Operating Systems Principles*. 441–459.
- [17] Dan Boneh. 1998. The Decision Diffie-Hellman Problem. In *International Algorithmic Number Theory Symposium*. Springer, 48–63.
- [18] Joseph Bonneau, Andrew Miller, Jeremy Clark, Arvind Narayanan, Joshua A. Kroll, and Edward W. Felten. 2015. Sok: Research Perspectives and Challenges for Bitcoin and Cryptocurrencies. In *2015 IEEE Symposium on Security and Privacy*. IEEE, 104–121.
- [19] Fabrice Boudot, Pierrick Gaudry, Aurore Guillevec, Nadia Heninger, Emmanuel Thomé, and Paul Zimmerman. 2020. Factorization of RSA-250. <https://lists.gforge.inria.fr/pipermail/cado-nfs-discuss/2020-February/001166.html>.
- [20] Justin Brickell and Vitaly Shmatikov. 2008. The Cost of Privacy: Destruction of Data-Mining Utility in Anonymized Data Publishing. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 70–78.
- [21] Daniel R.L. Brown and Kristian Gjøsteen. 2007. A Security Analysis of the NIST SP 800-90 Elliptic Curve Random Number Generator. In *Annual International Cryptology Conference*. Springer, 466–481.
- [22] Mark Bun, Damien Desfontaines, Cynthia Dwork, Moni Naor, Kobbi Nissim, Aaron Roth, Adam Smith, Thomas Steinke, Jonathan Ullman, and Salil Vadhan. 2021. Statistical Inference is Not a Privacy Violation. Posted June 3, 2021 at <https://differentialprivacy.org/inference-is-not-a-privacy-violation/>.
- [23] Joy Buolamwini and Timnit Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Conference on Fairness, Accountability and Transparency*. 77–91.
- [24] U.S. Census Bureau. 2021. Census Bureau Sets Key Parameters to Protect Privacy in 2020 Census Results. Posted on June 9th, 2021 at <https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance/2020-das-updates/2021-06-09.html>
- [25] U.S. Census Bureau. 2021. Disclosure Avoidance for the 2020 Census: An Introduction. U.S. Government Publishing Office, Washington, DC.
- [26] Ran Canetti, Oded Goldreich, and Shai Halevi. 2004. The Random Oracle Methodology, Revisited. *Journal of the ACM (JACM)* 51, 4 (2004), 557–594.
- [27] L. Elisa Celis, Damian Straszak, and Nisheeth K. Vishnoi. 2017. Ranking with Fairness Constraints. *arXiv preprint arXiv:1704.06840* (2017).
- [28] Shu-jeen Chang, Ray Perlner, William E. Burr, Meltem Sönmez Turan, John M. Kelsey, Souradyuti Paul, and Lawrence E. Bassham. 2012. *Third-Round Report of the SHA-3 Cryptographic Hash Algorithm Competition*. Technical Report. National Institute of Standards and Technology (NIST).
- [29] David Chaum. 1985. Security Without Identification: Transaction Systems to Make Big Brother Obsolete. *Commun. ACM* 28, 10 (1985), 1030–1044.
- [30] Albert Cheu, Adam Smith, Jonathan Ullman, David Zeber, and Maxim Zhilyaev. 2019. Distributed Differential Privacy via Shuffling. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 375–403.
- [31] Edmund M. Clarke, Thomas A. Henzinger, Helmut Veith, and Roderick Bloem. 2018. *Handbook of Model Checking*. Vol. 10. Springer.
- [32] Edmund M. Clarke Jr., Orna Grumberg, Daniel Kroening, Doron Peled, and Helmut Veith. 2018. *Model Checking*. MIT Press.
- [33] Véronique Cortier, David Galindo, Ralf Küsters, Johannes Mueller, and Tomasz Truderung. 2016. Sok: Verifiability Notions for E-Voting Protocols. In *2016 IEEE Symposium on Security and Privacy*. IEEE, 779–798.
- [34] George Danezis, Cedric Fournet, Markulf Kohlweiss, and Bryan Parno. 2013. Pinocchio Coin: Building Zerocoin from a Succinct Pairing-Based Proof System. In *Proceedings of the First ACM Workshop on Language Support for Privacy-Enhancing Technologies*. 27–30.
- [35] Damien Desfontaines and Balázs Pejó. 2020. SoK: Differential Privacies. *Proceedings on Privacy Enhancing Technologies* 2020, 2 (2020), 288–313.
- [36] Whitfield Diffie and Martin Hellman. 1976. New Directions in Cryptography. *IEEE Transactions on Information Theory* 22, 6 (1976), 644–654.
- [37] Whitfield Diffie and Susan Landau. 2007. The Export of Cryptography in the 20th Century and the 21st. In *The History of Information Security: A Comprehensive Handbook*, Karl De Leeuw and Jan Bergstra (Eds.). Elsevier, 725–736.
- [38] Irit Dinur and Kobbi Nissim. 2003. Revealing Information while Preserving Privacy. In *Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*. 202–210.

- [39] Danny Dolev and Andrew Yao. 1983. On the Security of Public Key Protocols. *IEEE Transactions on Information Theory* 29, 2 (1983), 198–208.
- [40] Finale Doshi-Velez and Been Kim. 2017. Towards a Rigorous Science of Interpretable Machine Learning. *arXiv preprint arXiv:1702.08608* (2017).
- [41] Cynthia Dwork. 2008. Differential Privacy: A Survey of Results. In *International Conference on Theory and Applications of Models of Computation*. Springer, 1–19.
- [42] Cynthia Dwork. 2009. The Differential Privacy Frontier. In *Theory of Cryptography Conference*. Springer, 496–502.
- [43] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. 2012. Fairness through Awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*. 214–226.
- [44] Cynthia Dwork, Nitin Kohli, and Deirdre Mulligan. 2019. Differential Privacy in Practice: Expose Your Epsilons! *Journal of Privacy and Confidentiality* 9, 2 (2019).
- [45] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. Calibrating Noise to Sensitivity in Private Data Analysis. In *Theory of Cryptography Conference*. Springer, 265–284.
- [46] Cynthia Dwork and Aaron Roth. 2014. The Algorithmic Foundations of Differential Privacy. *Foundations and Trends in Theoretical Computer Science* 9, 3-4 (2014), 211–407.
- [47] Cynthia Dwork and Adam Smith. 2010. Differential Privacy for Statistics: What We Know and What We Want to Learn. *Journal of Privacy and Confidentiality* 1, 2 (2010).
- [48] Úlfar Erlingsson, Vitaly Feldman, Ilya Mironov, Ananth Raghunathan, Kunal Talwar, and Abhradeep Thakurta. 2019. Amplification by Shuffling: From Local to Central Differential Privacy via Anonymity. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 2468–2479.
- [49] European Parliament and Council. 2016. Regulation (EU) 2016/679 of the European Parliament and of 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 (General Data Protection Regulation). *Official Journal of the European Union (OJ)* (2016).
- [50] Kevin Eykholt, Ivan Evtimov, Earlene Fernandes, Bo Li, Amir Rahmati, Chaowei Xiao, Atul Prakash, Tadayoshi Kohno, and Dawn Song. 2018. Robust Physical-World Attacks on Deep Learning Visual Classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '18)*. 1625–1634.
- [51] Amos Fiat and Adi Shamir. 1986. How to Prove Yourself: Practical Solutions to Identification and Signature Problems. In *Conference on the Theory and Application of Cryptographic Techniques*. Springer, 186–194.
- [52] Frank McSherry. February 25th, 2018. Uber’s differential privacy .. probably isn’t. <https://github.com/frankmcsherry/blog/blob/master/posts/2018-02-25.md>
- [53] Benjamin Fuller, Mayank Varia, Arkady Yerukhimovich, Emily Shen, Ariel Hamlin, Vijay Gadepally, Richard Shay, John Darby Mitchell, and Robert K. Cunningham. 2017. Sok: Cryptographically Protected Database Search. In *2017 IEEE Symposium on Security and Privacy*. IEEE, 172–191.
- [54] Simson L. Garfinkel, John M. Abowd, and Sarah Powazek. 2018. Issues Encountered Deploying Differential Privacy. In *Proceedings of the 2018 Workshop on Privacy in the Electronic Society*. 133–137.
- [55] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. 2018. Datasheets for Datasets. *arXiv preprint arXiv:1803.09010* (2018).
- [56] Oded Goldreich. 2006. On Post-Modern Cryptography. Cryptology ePrint Archive, Report 2006/461. <https://eprint.iacr.org/2006/461>.
- [57] Oded Goldreich. 2009. *Foundations of Cryptography: Volume 2, Basic Applications*. Cambridge University Press.
- [58] Shafi Goldwasser and Yael Tauman Kalai. 2003. On the (In)security of the Fiat-Shamir Paradigm. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science (FOCS '03)*. IEEE, 102–113.
- [59] Shafi Goldwasser and Yael Tauman Kalai. 2016. Cryptographic Assumptions: A Position Paper. In *Theory of Cryptography Conference*. Springer, 505–522.
- [60] Shafi Goldwasser and Silvio Micali. 1984. Probabilistic Encryption. *J. Comput. System Sci.* 28, 2 (1984), 270–299.
- [61] Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. 2014. Explaining and Harnessing Adversarial Examples. *arXiv preprint arXiv:1412.6572* (2014).
- [62] <https://blog.cryptographyengineering.com/2013/09/18/the-many-flaws-of-dualecdrbg/>.
- [63] Jens Groth. 2010. Short Pairing-based Non-Interactive Zero-Knowledge Arguments. In *International Conference on the Theory and Application of Cryptology and Information Security*. Springer, 321–340.
- [64] Gus Wezerek and David Van Riper. February 6th, 2020. Changes to the Census Could Make Small Towns Disappear. *The New York Times Opinion Section* (February 6th, 2020). <https://www.nytimes.com/interactive/2020/02/06/opinion/census-algorithm-privacy.html>.
- [65] J. Alex Halderman, Seth D. Schoen, Nadia Heninger, William Clarkson, William Paul, Joseph A. Calandrino, Ariel J. Feldman, Jacob Appelbaum, and Edward W. Felten. 2009. Lest We Remember: Cold-Boot Attacks on Encryption Keys. *Commun. ACM* 52, 5 (2009), 91–98.
- [66] Nadia Heninger, Zakir Durumeric, Eric Wustrow, and J. Alex Halderman. 2012. Mining your Ps and Qs: Detection of Widespread Weak Keys in Network Devices. In *21st USENIX Security Symposium (USENIX Security '12)*. 205–220.
- [67] Justin Hsu, Marco Gaboardi, Andreas Haeberlen, Sanjeev Khanna, Arjun Narayan, Benjamin C. Pierce, and Aaron Roth. 2014. Differential Privacy: An Economic Method for Choosing Epsilon. In *2014 IEEE 27th Computer Security Foundations Symposium*. IEEE, 398–410.
- [68] Jessica Hullman. 2021. Shots taken, shots returned regarding the Census’ motivation for using differential privacy (and btw, it’s not an algorithm). Posted August 27, 2021 at <https://statmodeling.stat.columbia.edu/2021/08/27/shots-taken-shots-returned-regarding-the-census-motivation-for-using-differential-privacy-and-btw-its-not->

- [69] Russell Impagliazzo. 1995. A Personal View of Average-Case Complexity. In *Proceedings of the Tenth Annual IEEE Conference on Structure in Complexity Theory*. IEEE, 134–147.
- [70] Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. 2011. What Can We Learn Privately? *SIAM J. Comput.* 40, 3 (2011), 793–826.
- [71] Jonathan Katz and Yehuda Lindell. 2014. *Introduction to Modern Cryptography*. CRC Press.
- [72] Christopher T. Kenny, Shiro Kuriwaki, Cory McCartan, Evan Rosenman, Tyler Simko, and Kosuke Imai. 2021. The Impact of the US Census Disclosure Avoidance System on Redistricting and Voting Rights Analysis. *arXiv preprint arXiv:2105.14197* (2021).
- [73] Neal Koblitz and Alfred J. Menezes. 2007. Another Look at "Provable Security". *Journal of Cryptology* 20, 1 (2007), 3–37.
- [74] Nitin Kohli and Paul Laskowski. 2018. Epsilon Voting: Mechanism Design for Parameter Selection in Differential Privacy. In *2018 IEEE Symposium on Privacy-Aware Computing (PAC)*. IEEE, 19–30.
- [75] Sara Krehbiel. 2019. Choosing Epsilon for Privacy as a Service. *Proceedings on Privacy Enhancing Technologies* 2019, 1 (2019), 192–205.
- [76] Joshua A. Kroll, Solon Barocas, Edward W. Felten, Joel R. Reidenberg, David G. Robinson, and Harlan Yu. 2016. Accountable Algorithms. *U. Pa. L. Rev.* 165 (2016), 633.
- [77] Peeter Laud and Alisa Pankova. 2019. Interpreting Epsilon of Differential Privacy in Terms of Advantage in Guessing or Approximating Sensitive Attributes. *arXiv preprint arXiv:1911.12777* (2019).
- [78] Jaewoo Lee and Chris Clifton. 2011. How Much is Enough? Choosing  $\epsilon$  for Differential Privacy. In *International Conference on Information Security*. Springer, 325–340.
- [79] Lawrence Lessig. 1999. *Code and Other Laws of Cyberspace*. Basic Books.
- [80] Changchang Liu, Xi He, Thee Chanyaswad, Shiqiang Wang, and Prateek Mittal. 2019. Investigating Statistical Privacy Frameworks from the Perspective of Hypothesis Testing. *Proceedings on Privacy Enhancing Technologies* 2019, 3 (2019), 233–254.
- [81] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. 2017. Towards Deep Learning Models Resistant to Adversarial Attacks. *arXiv preprint arXiv:1706.06083* (2017).
- [82] Momin M. Malik. 2020. A Hierarchy of Limitations in Machine Learning. *arXiv preprint arXiv:2002.05193* (2020).
- [83] Marcela S. Melara, Aaron Blankstein, Joseph Bonneau, Edward W. Felten, and Michael J. Freedman. 2015. CONIKS: Bringing Key Transparency to End Users. In *24th USENIX Security Symposium (USENIX Security '15)*. 383–398.
- [84] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model Cards for Model Reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*. 220–229.
- [85] Deirdre K. Mulligan, Joshua A. Kroll, Nitin Kohli, and Richmond Y. Wong. 2019. This Thing Called Fairness: Disciplinary Confusion Realizing a Value in Technology. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–36.
- [86] Moni Naor. 2003. On Cryptographic Assumptions and Challenges. In *Annual International Cryptology Conference*. Springer, 96–109.
- [87] Arvind Narayanan and Vitaly Shmatikov. 2008. Robust de-Anonymization of Large Sparse Datasets. In *2008 IEEE Symposium on Security and Privacy (IEEE S&P '08)*. IEEE, 111–125.
- [88] Helen Nissenbaum. 1996. Accountability in a Computerized Society. *Science and engineering Ethics* 2, 1 (1996), 25–42.
- [89] Safiya Umoja Noble. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press.
- [90] Paul Ohm. 2009. Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. *UCLA L. Rev.* 57 (2009), 1701.
- [91] Nicolas Papernot, Patrick McDaniel, Arunesh Sinha, and Michael P. Wellman. 2018. SoK: Security and Privacy in Machine Learning. In *2018 IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE, 399–414.
- [92] Bryan Parno, Jon Howell, Craig Gentry, and Mariana Raykova. 2013. Pinocchio: Nearly Practical Verifiable Computation. In *2013 IEEE Symposium on Security and Privacy*. IEEE, 238–252.
- [93] Balázs Pejó, Qiang Tang, and Gergely Biczók. 2019. Together or Alone: The Price of Privacy in Collaborative Learning. *Proceedings on Privacy Enhancing Technologies* 2019, 2 (2019), 47–65.
- [94] Colin Percival. 2005. Cache Missing for Fun and Profit. Available at <http://css.csail.mit.edu/6.858/2014/readings/ht-cache.pdf>.
- [95] Trevor Perrin. 2016. *The XEdDSA and VXEdDSA Signature Schemes*. Technical Report. Signal Foundation.
- [96] Oded Regev. 2009. On Lattices, Learning with Errors, Random Linear Codes, and Cryptography. *Journal of the ACM (JACM)* 56, 6 (2009), 1–40.
- [97] Ronald L. Rivest, Adi Shamir, and Leonard Adleman. 1978. A Method for Obtaining Digital Signatures and Public-Key Cryptosystems. *Commun. ACM* 21, 2 (1978), 120–126.
- [98] Phillip Rogaway. 2015. The Moral Character of Cryptographic Work. Cryptology ePrint Archive, Report 2015/1162. <https://eprint.iacr.org/2015/1162>.
- [99] Cynthia Rudin. 2019. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nature Machine Intelligence* 1, 5 (2019), 206–215.
- [100] Steven Ruggles and David Van Riper. 2021. The Role of Chance in the Census Bureau Database Reconstruction Experiment. *Population Research and Policy Review* (2021).
- [101] Alexis R. Santos-Lozada, Jeffrey T. Howard, and Ashton M. Verdery. 2020. How Differential Privacy Will Affect our Understanding of Health Disparities in the United States. *Proceedings of the National Academy of Sciences* (2020).

- [102] Sarah Scheffler and Mayank Varia. 2021. Protecting Cryptography Against Compelled Self-Incrimination. In *30th USENIX Security Symposium (USENIX Security '21)*. To appear, available at [https://www.usenix.org/system/files/sec21summer\\_scheffler.pdf](https://www.usenix.org/system/files/sec21summer_scheffler.pdf).
- [103] Andrew D. Selbst, danah boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet Vertesi. 2019. Fairness and Abstraction in Sociotechnical Systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*. 59–68.
- [104] Shai Shalev-Shwartz and Shai Ben-David. 2014. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press.
- [105] R.L. Sumwalt, B. Landsberg, and J. Homendy. 2019. *Assumptions Used in the Safety Assessment Process and the Effects of Multiple Alerts and Indications on Pilot Performance*. Technical Report. National Transportation Safety Board. <https://www.ntsb.gov/investigations/AccidentReports/Reports/ASR1901.pdf>.
- [106] Jun Tang, Aleksandra Korolova, Xiaolong Bai, Xueqiang Wang, and Xiaofeng Wang. 2017. Privacy Loss in Apple’s Implementation of Differential Privacy on MacOS 10.12. *arXiv preprint arXiv:1709.02753* (2017).
- [107] Kurt Thomas, Devdatta Akhawe, Michael Bailey, Dan Boneh, Elie Bursztein, Sunny Consolvo, Nicola Dell, Zakir Durumeric, Patrick Gage Kelley, Deepak Kumar, et al. 2021. SoK: Hate, Harassment, and the Changing Landscape of Online Abuse. (2021).
- [108] Dimitris Tsipras, Shibani Santurkar, Logan Engstrom, Alexander Turner, and Aleksander Madry. 2018. Robustness May be at Odds with Accuracy. *arXiv preprint arXiv:1805.12152* (2018).
- [109] Nik Unger, Sergej Dechand, Joseph Bonneau, Sascha Fahl, Henning Perl, Ian Goldberg, and Matthew Smith. 2015. SoK: Secure Messaging. In *2015 IEEE Symposium on Security and Privacy*. IEEE, 232–249.
- [110] Larry Wasserman and Shuheng Zhou. 2010. A Statistical Framework for Differential Privacy. *J. Amer. Statist. Assoc.* 105, 489 (2010), 375–389.
- [111] Daniel J. Weitzner, Harold Abelson, Tim Berners-Lee, Joan Feigenbaum, James Hendler, and Gerald Jay Sussman. 2008. Information Accountability. *Commun. ACM* 51, 6 (2008), 82–87.
- [112] Alexandra Wood, Micah Altman, Aaron Bembek, Mark Bun, Marco Gaboardi, James Honaker, Kobbi Nissim, David R. O’Brien, Thomas Steinke, and Salil Vadhan. 2018. Differential Privacy: A Primer for a Non-Technical Audience. *Vanderbilt Journal of Entertainment & Technology Law* 21, 1 (2018), 209–275.
- [113] Aiping Xiong, Tianhao Wang, Ninghui Li, and Somesh Jha. 2020. Towards Effective Differential Privacy Communication for Users’ Data Sharing Decision and Comprehension. *arXiv preprint arXiv:2003.13922* (2020).