Differential Privacy as a Causal Property

Michael Tschantz[§], Shayak Sen[*], Anupam Datta[*]

TR-17-001

October 2017

Abstract

We present associative and causal views of differential privacy. Under the associative view, the possibility of dependencies between data points precludes a simple statement of differential privacy's guarantee as conditioning upon a single changed data point. However, a simple characterization of differential privacy as limiting the effect of a single data point does exist under the causal view, without independence assumptions about data points. We believe this characterization resolves disagreement and confusion in prior work about the consequences of differential privacy. It also opens up the possibility of applying results from statistics, experimental design, and science about causation while studying differential privacy.

§ International Computer Science Institute, 1947 Center Street, Suite 600, Berkeley, California, 94704
* Carnegie Mellon University 5000 Forbes Ave, Pittsburg, Pennsylvania, 15213

Differential Privacy as a Causal Property

Michael Carl Tschantz, International Computer Science Institute
Shayak Sen, Carnegie Mellon University
Anupam Datta, Carnegie Mellon University

October 17, 2017

We present associative and causal views of differential privacy. Under the associative view, the possibility of dependencies between data points precludes a simple statement of differential privacy's guarantee as conditioning upon a single changed data point. However, a simple characterization of differential privacy as limiting the effect of a single data point does exist under the causal view, without independence assumptions about data points. We believe this characterization resolves disagreement and confusion in prior work about the consequences of differential privacy. It also opens up the possibility of applying results from statistics, experimental design, and science about causation while studying differential privacy.

1. Introduction

Differential privacy is a precise mathematical property of an algorithm requiring that it produce almost identical distributions of outputs for any pair of possible input databases that differs in a single data point. A disagreement has arisen in the literature with some researchers feeling that differential privacy makes an implicit assumption of independence between data points (e.g., [1, 2, 3, 4, 5]) and others asserting that no such assumption exists (e.g., [6, 7, 8, 9]). How can such a disagreement arise about a precise mathematical property of an algorithm?

We believe that the disagreement is not actually about differential privacy itself but rather about the meaning of an intuitive consequence of differential privacy commonly used to explain why it protects privacy. Kasivisiwanathan and Smith express this intuition as follows [7]:

> This definition states that changing a single individual's data in the database leads to a small change in the distribution on outputs.

This consequence of differential privacy, used to provide an intuitive characterization of it, does not make explicit the notion of *change* intended. In more detail, the above intuitive sentence compares the distribution over the output, a random variable O, in two hypothetical worlds, the pre- and post-change worlds. If we let D_i be a random variable representing the changed data point and d_i and d_i' be the pre- and post-change values for D_i, then the comparison is between $\Pr[O{=}o \text{ when } D_i{=}d_i]$ and $\Pr[O{=}o \text{ when } D_i{=}d_i']$. The part of this characterization of differential privacy that is informal is the notion of *when*, which would make the notion of *change* precise.

This paper contrasts two interpretations of changing inputs and *when*. The first we consider is *conditioning* upon two different values for the changed data

Num.	Conditions on $\mathcal{P}$	Point of comparison	Relation	Appears in
Original Differential Privacy				
1		$\Pr_{\mathcal{A}}[\mathcal{A}(d_1, \ldots, \underline{d_i^{\cdot}}, \ldots, d_n){=}o]$	is dp	[10]
Associative Variants				
4	$\Pr_{\mathcal{P}}[D_1{=}d_1, \ldots, D_i{=}\underline{d_i^{\cdot}}, \ldots, D_n{=}d_n] > 0$	$\Pr_{\mathcal{P}}[O{=}o \mid D_1{=}d_1, \ldots, D_i{=}\underline{d_i^{\cdot}}, \ldots, D_n{=}d_n]$	$\leftarrow$ dp	
6	$\Pr_{\mathcal{P}}[D_i{=}d_i] > 0$	$\Pr_{\mathcal{P}}[O{=}o \mid D_i{=}\underline{d_i^{\cdot}}]$	none	[1]
8	indep. D_i, $\Pr_{\mathcal{P}}[D_i{=}d_i] > 0$	$\Pr_{\mathcal{P}}[O{=}o \mid D_i{=}\underline{d_i^{\cdot}}]$	$\leftarrow$ dp	[1]
Causal Variants				
10		$\Pr_{\mathcal{P}}[O{=}o \mid \mathrm{do}(D_1{=}d_1, \ldots, D_i{=}\underline{d_i^{\cdot}}, \ldots, D_n{=}d_n)]$	$\leftrightarrow$ dp	
12		$\Pr_{\mathcal{P}}[O{=}o \mid \mathrm{do}(D_i{=}\underline{d_i^{\cdot}})]$	$\leftarrow$ dp	

Table 1: Various definitions similar to Differential Privacy. The left-most column gives the number used to identify the definition, where they are numbered by the order in which they appear later in the text. The propositions are numbered such that the number identifying a definition also identifies the proposition showing that definition's relationship with differential privacy. The point of comparison is the quantity computed twice, once for two different values of the ith data point, and compared to check whether they are within a factor of e^{ϵ} of one another. The check is for all values of the index i and all pairs of data values d_i and d_i' that can go in $\underline{d_i^{\cdot}}$. In one case (Definition 8), the comparison just applies to distributions where the data points are independent of one another. Some of definitions only perform the comparison when changed data point D_i having the value d_i (and d_i', the changed value) has non-zero probability under $\mathcal{P}$. Others only perform the comparison when all the data points D having the values d (for original and changed value of d_i) has non-zero probability. do denotes a causal intervention instead of standard conditioning [11].

point. This interpretation focuses on two different subsets of an input space and accounts for associations between data points in the database. This associative interpretation captures what a rational agent would do upon seeing one or the other input value in a natural, observational setting. Furthermore, as we will discuss in more detail below, the associative view turns out to match up with the views of those believing differential privacy has an implicit assumption of independence, that is, a lack of association.

The second interpretation we consider is *intervention* in a causal model. This interpretation models artificially altering inputs, as in an experiment. While it tracks causal effects by accounting for how the intervention may cause other values to change, it ignores associations in the database since such artificial interventions break them. As such, the purported implicit assumption disappears and this interpretation more tightly characterizes the consequences of differential privacy.

Table 1 provides an overview of our results about various interpretations of the key consequence of differential privacy quoted above as an intuitive characterization of it. After reviewing differential privacy (Section 3), we start our analysis with the associative view using conditioning. We first consider conditioning upon all the data points instead of just the changed one (Section 4). After dealing with some annoyances involving the inability to condition on zero-probability data points, we get a precise characterization of differential privacy's consequences (Proposition 4). However, this associative definition does not correspond well to the intuitive characterization of differential privacy's key consequences quoted above: whereas the above quoted characterization refers to just the changed data point, the associative definition refers to them all blurring the characterization's focus on change.

We next consider conditioning upon just the single changed data point (Sec-

tion 5). Doing so produces a stronger definition not implied by differential privacy (Proposition 6). The definition is, however, implied with the additional assumption of independence between data points (Proposition 8). We believe this explains the feeling some have that differential privacy implicitly assumes such: to get the key characterization of differential privacy's consequences to hold appears to require such an assumption.

However, we go on to show that the assumption is not required when using a causal interpretation of the key consequence of differential privacy quoted above. As a warm-up exercise, we first consider intervening upon all the data points after reviewing the key concepts of causal modeling (Section 6). As before, referring to all data points produces a definition that characterizes differential privacy but without the intuitive focus on a single data point we desire (Proposition 10).

We then consider intervening upon a single point (Section 7). We find that this causal characterization of differential privacy is in fact implied by differential privacy without any assumptions about independence (Proposition 12). An additional benefit we find is that, unlike the associative characterizations, we need no side conditions limiting the characterization to data points with non-zero probabilities. This benefit follows from causal interventions being defined for zero-probability events unlike conditioning upon them. For these two reasons, we believe that differential privacy is better viewed as a causal property than as an associative one.

In addition to considering the consequences of differential privacy through the lenses of association and causation, we also consider how these two approaches can provide definitions equivalent to differential privacy. Table 2 shows our key results about definitions that are either equivalent to differential privacy or might be mistaken as such, which, in the sections below, we weave in with our aforementioned results about characterizations of the consequences of differential privacy.

When intervening upon all data points, we get equivalence for free from Definition 10 that we already explored as a characterization of the consequences of differential privacy. This free equivalence does not occur for conditioning upon all data points since the side condition ruling out zero-probability data points means those data points are not constrained by Definition 4. Since differential privacy is a restriction on all data points, to get an equivalence, the definition must check all data points. To achieve this, we further require that the definition hold on all distributions over the data points, not just the naturally occurring one. (Alternatively, we could require the definition to hold for any one distribution with non-zero probabilities for all data points, such as the uniform distribution.) We also make similar alterations to the definitions looking a single data point.

As we elaborate in the conclusion (Section 8), these results open up the possibility of using all the methods developed for working with causation to work with differential privacy. Furthermore, we show that the difference between

Num.	$\mathcal{P}$	Conditions	Point of comparison	Relation
Original Differential Privacy				
1			$\Pr_{\mathcal{A}}[\mathcal{A}(d_1, \ldots, \underline{d_i^{\cdot}}, \ldots, d_n){=}o]$	is dp
Associative Variants				
5	$\forall$	$\Pr_{\mathcal{P}}[D_1{=}d_1, \ldots, D_i{=}\underline{d_i^{\cdot}}, \ldots, D_n{=}d_n] > 0$	$\Pr_{\mathcal{P}}[O{=}o \mid D_1{=}d_1, \ldots, D_i{=}\underline{d_i^{\cdot}}, \ldots, D_n{=}d_n]$	$\leftrightarrow$ dp
7	$\forall$	$\Pr_{\mathcal{P}}[D_i{=}d_i] > 0$	$\Pr_{\mathcal{P}}[O{=}o \mid D_i{=}\underline{d_i^{\cdot}}]$	$\rightarrow$ dp
9	$\forall$ indep. D_i	$\Pr_{\mathcal{P}}[D_i{=}d_i] > 0$	$\Pr_{\mathcal{P}}[O{=}o \mid D_i{=}\underline{d_i^{\cdot}}]$	$\leftrightarrow$ dp
Causal Variants				
10			$\Pr_{\mathcal{P}}[O{=}o \mid \mathrm{do}(D_1{=}d_1, \ldots, D_i{=}\underline{d_i^{\cdot}}, \ldots, D_n{=}d_n)]$	$\leftrightarrow$ dp
11	$\forall$		$\Pr_{\mathcal{P}}[O{=}o \mid \mathrm{do}(D_1{=}d_1, \ldots, D_i{=}\underline{d_i^{\cdot}}, \ldots, D_n{=}d_n)]$	$\leftrightarrow$ dp
13	$\forall$		$\Pr_{\mathcal{P}}[O{=}o \mid \mathrm{do}(D_i{=}\underline{d_i^{\cdot}})]$	$\leftrightarrow$ dp

Table 2: Various definitions similar to Differential Privacy. The notation is the same as in Table 1. The definitions vary in whether they require performing these comparisons for just the actual probability distribution over data points $\mathcal{P}$ or over all such distributions. In one case (Definition 9), the comparison just applies to distributions where the data points are independent of one another.

the two views of differential privacy is precisely captured as the difference between association and causation. That some fail to get what they want out of differential privacy (without making an unrealistic assumption of independence) comes from the contrapositive of the maxim *correlation doesn't imply causation*: differential privacy ensuring a lack of (strong) causation does not imply a lack of (strong) association. Given the common confusion of association and causation, and that differential privacy does not make its causal nature explicit in its mathematical statement, we believe our work explains how reasonable researchers can be in apparent disagreement about the meaning (really, consequences) of differential privacy.

2. Prior Work

The paper coining the term "differential privacy" recognized that causation was key to understanding differential privacy: "it will not be the presence of her data that causes [the disclosure of sensitive information]" [12, page 8]. Despite this causal view being present in the understanding of differential privacy from the beginning, we believe we are first to make it precise and to compare it explicitly with an associative view.

Kasivisiwanathan and Smith look at a different way of comparing the two views of differential privacy [7]. They study the Bayesian probabilities that an adversary would assign, after seeing the system's outputs, to a property holding of a data provider. They compare these probabilities under various possible inputs that a data provider could provide. For systems with differential privacy, they show that the Bayesian probabilities hardly change under the different inputs. This provides a Bayesian interpretation of differential privacy

without making an assumption of independent data points. Kasivisiwanathan and Smith also comment that such an assumption would be required when comparing Bayesian probabilities before and after seeing the system's output. We instead work with only physical or frequentist probabilities and instead find a difference between association and causation.

This work is largely motivated by wanting to explain the difference between two camps that have emerged around differential privacy. The first camp, associated with the inventors of differential privacy, emphasizes differential privacy's ability to ensure that data providers are no worse off for providing data (e.g., [12, 7, 8, 9]). The second camp, which formed in response to limitations in differential privacy's guarantee, emphasizes that an adversary should not be able to learn anything sensitive about the data providers after the system releases outputs computed from data from data providers (e.g., [1, 2, 3, 4, 5]). The second camp notes that differential privacy fails to provide this guarantee when the data points from different data providers are associated with one another. McSherry provides an informal description of the disagreement between the camps [8].

We provide a mathematically precise characterization of what each camp wants and an explanation of how two camps can grow up around the precise mathematical definition of differential privacy. Noting that the second camp expresses their desires for privacy in terms of association and conditional probabilities common to information theory and quantitative information flow (see Smith [13] for a survey), we start by attempting to express differential privacy in such terms. A clean expression of differential privacy in terms of conditioning upon a single participant's data point only emerges in cases where data points are not associated with one another. This result explains the essence of the second camp's complaint that "differential privacy mechanisms assume independence of tuples [i.e., data points] in the database" [5, page 1].

However, we find that the purported assumption is not required to precisely state differential privacy in terms of causation, where conditioning upon the data point is replaced by causally intervening upon it. This causal characterization justifies the first camp's rebuttal that differential privacy provides a different but meaningful guarantee from the one expected by the second camp.

While not necessary for understanding our technical development, Appendix A provides a history of the two competing views of differential privacy.

3. Differential Privacy

Kasivisiwanathan and Smith restate the definition differential privacy as follows [7]:

> Databases are assumed to be vectors in $\mathcal{D}^n$ for some domain $\mathcal{D}$. The Hamming distance $d_H(\vec{x}, \vec{y})$ on $\mathcal{D}^n$ is the number of positions in which the vectors $\vec{x}, \vec{y}$ differ. We let $\Pr[\cdot\,]$ and $\mathbb{E}[\cdot\,]$ denote probability and expectation, respectively. Given a randomized algorithm $\mathcal{A}$, we let $\mathcal{A}(\vec{x})$ be the random variable (or, probability

distribution on outputs) corresponding to input $\vec{x}$. [...]

Definition 1.1 (ϵ-differential privacy [7]). *A randomized algorithm $\mathcal{A}$ is said to be ϵ-differentially private if for all databases $\vec{x}, \vec{y} \in \mathcal{D}^n$ at Hamming distance at most 1, and for all subsets S of outputs,*

$$\Pr[\mathcal{A}(\vec{x}) \in S] \leq e^{\epsilon} \cdot \Pr[\mathcal{A}(\vec{y}) \in S].$$

This definition states that changing a single individual's data in the database leads to a small change in the distribution on outputs.

(The reference "[7]" in this quote refers to Dwork et al.'s paper, which we refer to as [10], and not to the reference numbered 7 in the paper you are currently reading.) For simplicity, we will limit our discussion to the discrete case, in which checking for membership in a set of outputs can be replaced with checking for equality to a particular output. We further simplify by limiting ourselves to the considering data points that range over a finite set $\mathcal{D}$ and the outputs that range over a finite set $\mathcal{O}$. We also rename some of the variables.

Definition 1. A randomized algorithm $\mathcal{A}$ is said to be *ϵ-differentially private (in the discrete case)* if for all databases $d, d' \in \mathcal{D}^n$ at Hamming distance at most 1, and for all output values o,

$$\Pr_{\mathcal{A}}[\mathcal{A}(d)=o] \leq e^{\epsilon} * \Pr_{\mathcal{A}}[\mathcal{A}(d')=o] \tag{1}$$

The probabilities are frequencies that refer to unpredictable and independent randomization in the algorithm $\mathcal{A}$. The probabilities do not depend on anything like the distribution over the databases d or d', which are values, not random variables, taken as provided as inputs. We remind us of this, we subscripted $\Pr$ with $\mathcal{A}$ to make explicit what the frequencies are over, but we will drop it when there is no risk of confusion.

These two definitions are mathematically precise conditions on the algorithm $\mathcal{A}$. However, going from these conditions to the intuition captured by the last quoted sentence about changing data is not as transparent as it could be.

First, it refers to "the database" but where is "the database" represented in these definitions? In a sense it's d and d', but then there's two of them. Rather, "the database" appears to refer to the formal argument of $\mathcal{A}$, which is unseen. By not having the database explicitly named, it is difficult to precisely discuss changes to it. To make things more explicit, let us name the database D. Since the database can take on more than one value, D is a random variable. Much as d and d' are vectors of values, the random variable D ranges over vectors of values. Let D_i be a random variable over the ith such value, that is, the input from the ith individual in the database. D is related to $D_1, \ldots, D_n$ informally as $D = \langle D_1, \ldots, D_n \rangle$ and more formally as $D(\omega) = \langle D_1(\omega), \ldots, D_n(\omega) \rangle$ where ω ranges over the outcome space of the probability space. Either way, $\Pr_{\mathcal{P}}[D{=}d] = \Pr_{\mathcal{P}}[\langle D_1, \ldots, D_n \rangle{=}d]$ for all value vectors d representing databases. Here, we subscripted $\Pr$ with $\mathcal{P}$

instead of $\mathcal{A}$ because the data points come from some population $\mathcal{P}$ of individuals that determines their frequencies and these frequencies are independent of the randomization within $\mathcal{A}$. Note, however, that these frequencies are irrelevant to the definition of differential privacy since it only refers the frequencies produced by the randomization within the algorithm $\mathcal{A}$.

Second, the above quote refers to "the distribution on outputs". Typically, we think of random variables as having distributions, leading to the question of which random variable is the output random variable. As before, the obvious answer of $\mathcal{A}(d)$ and $\mathcal{A}(d')$ leads to two random variables instead of one. So, we react similarly and introduce an explicit name O for the output and treat that as the single random variable where informally $O = \mathcal{A}(D)$, or more formally, $O(\omega) = \mathcal{A}(D(\omega))(\omega)$, where $D(\omega)$ denotes the value that D takes in outcome ω and $\mathcal{A}(d)(\omega)$ denotes the output of $\mathcal{A}$ when given the input database d and its randomization is resolved by ω. That is, $\Pr_{\mathcal{P},\mathcal{A}}[O{=}o] = \Pr_{\mathcal{P},\mathcal{A}}[\mathcal{A}(D){=}o]$ where the frequencies depend upon both the population $\mathcal{P}$ and algorithm $\mathcal{A}$. Since $O = \mathcal{A}(D_1, \ldots, D_n)$ and the internal randomness of $\mathcal{A}$ is independent of D, $\Pr_{\mathcal{P}}[D_1{=}d_1, \ldots, D_n{=}d_n] > 0$ implies

$$\Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid D_1{=}d_1, \ldots, D_n{=}d_n] = \Pr_{\mathcal{A}}[\mathcal{A}(d_1, \ldots, d_n){=}o] \qquad (2)$$

for all populations $\mathcal{P}$.

Using these explicit random variables we can restate the above quoted characterization of the consequences of differential privacy as

> This definition states that changing the value of a single D_i in the database D leads to a small change in the distribution on outputs O.

Let us similarly restate the definition of differential privacy to make the database explicit. An almost formal attempt might be

Definition 2 (undefined)**.** A randomized algorithm $\mathcal{A}$ is said to be *ϵ-differentially private with an undefined when* if for all databases $d, d' \in \mathcal{D}^n$ at Hamming distance at most 1, and for all output values o,

$$\Pr[O{=}o \text{ when } D{=}d] \leq e^{\epsilon} * \Pr[O{=}o \text{ when } D{=}d'] \qquad (3)$$

where $O = \mathcal{A}(D)$ and $D = \langle D_1, \ldots, D_n \rangle$.

Here, the problem is that "when" is not precisely defined.

4. Differential Privacy as Association with the Whole Database

The obvious way to make "when" precise is with conditioning. We can attempt to define differential privacy in terms of a comparison of two conditional probabilities where the difference between them is a difference in the conditioned upon value.

Definition 3 (sometimes undefined)**.** A randomized algorithm $\mathcal{A}$ is said to be *ϵ-differentially private as conditioning on the whole database* if for all

databases $d, d' \in \mathcal{D}^n$ at Hamming distance at most 1, and for all output values o,

$$\Pr[O{=}o \mid D{=}d] \leq e^{\epsilon} * \Pr[O{=}o \mid D{=}d'] \tag{4}$$

where $O = \mathcal{A}(D)$ and $D = \langle D_1, \ldots, D_n \rangle$.

This definition is not equivalent to Definition 1 because the conditional probabilities referenced are not defined whenever $\Pr[D{=}d] = 0$ since $\Pr[O{=}o \mid D{=}d] = \Pr[O{=}o \wedge D{=}d] / \Pr[D{=}d]$. (And the same goes $D = d'$.) That is, the definition of differential privacy considers databases that might not occur naturally, but conditioning upon them is undefined.

To avoid this issue, one can restrict his attention to data points with non-zero probabilities:

Definition 4 (implied, but weaker)**.** A randomized algorithm $\mathcal{A}$ is said to be *ϵ-differentially private as qualified conditioning on the whole database* if for all databases $d, d' \in \mathcal{D}^n$ at Hamming distance at most 1, and for all output values o, if $\Pr[D{=}d] > 0$ and $\Pr[D{=}d'] > 0$ then

$$\Pr[O{=}o \mid D{=}d] \leq e^{\epsilon} * \Pr[O{=}o \mid D{=}d'] \tag{5}$$

where $O = \mathcal{A}(D)$ and $D = \langle D_1, \ldots, D_n \rangle$.

Definition 4 is implied by differential privacy but is weaker than it, making it a characterization of differential privacy's consequences. It is weaker since it places no requirements on the behavior of $\mathcal{A}$ for inputs with zero probability. By being a property about a algorithm operating on a single fixed distribution over data points, the actual distribution occurring in practice, such zero-probability data points will exist whenever nature constrains the values that data points can take on.

Since the definition is only weaker on zero-probability inputs, this change might seem unimportant. However, it introduces possible information leaks whenever the adversary does not realize that a particular input has zero probability. For example, suppose $\Pr[D_i{=}2] = 0$. The behavior of $\mathcal{A}$ given D_i with the value 2 is unconstrained by Definition 4 and it might never produce an output $o_{\neg 2}$ that it otherwise produces with non-zero probability. Then, an adversary will, upon not seeing $o_{\neg 2}$ will learn that D_i was not 2. If the adversary did not know that $\Pr[D_i{=}2] = 0$, this will be new information for the adversary.

(We start numbering propositions from 4 to align their numbering with that of the definitions about which they are.)

Proposition 4. Definition 1 implies Definition 4, but not the other way around.

Proof. Assume Definition 1 holds. Consider any population $\mathcal{P}$, index i, data points $d_1, \ldots, d_n$ in $\mathcal{D}^n$ and d'_i in $\mathcal{D}$, and output o such that the following hold: $\Pr_{\mathcal{P}}[D_1{=}d_1, \ldots, D_n{=}d_n] > 0$ and $\Pr_{\mathcal{P}}[D_1{=}d_1, \ldots, D_i{=}d'_i, \ldots, D_n{=}d_n] > 0$. Since Definition 1 holds,

$$\Pr_{\mathcal{A}}[\mathcal{A}(d_1,\ldots,d_n){=}o] \leq e^{\epsilon} * \Pr_{\mathcal{A}}[\mathcal{A}(d_1,\ldots,d_i,\ldots,d_n){=}o] \tag{6}$$

$$\Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid D_1{=}d_1,\ldots,D_n{=}d_n] \leq e^{\epsilon} * \Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid D_1{=}d_1,\ldots,D_i{=}d_i',\ldots,D_n{=}d_n] \tag{7}$$

where the second line follows from (2). Thus, Definition 4 holds.

To prove that Definition 4 does not imply Definition 1, consider the case of a database holding a single data point whose value could be 0, 1, or 2. Suppose the population $\mathcal{P}$ is such that $\Pr_{\mathcal{P}}[D_1{=}2] = 0$. Consider an algorithm $\mathcal{A}$ such that for the given population $\mathcal{P}$,

$$\Pr_{\mathcal{A}}[\mathcal{A}(0){=}0] = 1/2 \qquad \Pr_{\mathcal{A}}[\mathcal{A}(0){=}1] = 1/2 \tag{8}$$

$$\Pr_{\mathcal{A}}[\mathcal{A}(1){=}0] = 1/2 \qquad \Pr_{\mathcal{A}}[\mathcal{A}(1){=}1] = 1/2 \tag{9}$$

$$\Pr_{\mathcal{A}}[\mathcal{A}(2){=}0] = 1 \qquad \Pr_{\mathcal{A}}[\mathcal{A}(2){=}1] = 0 \tag{10}$$

The algorithm does not satisfy Definition 1 due to its behavior on the input 2. However, using (2),

$$\Pr_{\mathcal{P},\mathcal{A}}[O{=}0 \mid D_1{=}0] = 1/2 \qquad \Pr_{\mathcal{P},\mathcal{A}}[O{=}1 \mid D_1{=}0] = 1/2 \tag{11}$$

$$\Pr_{\mathcal{P},\mathcal{A}}[O{=}0 \mid D_1{=}1] = 1/2 \qquad \Pr_{\mathcal{P},\mathcal{A}}[O{=}1 \mid D_1{=}1] = 1/2 \tag{12}$$

While (2) says nothing about $D_1{=}2$ since that has zero probability, this is sufficient to show that the algorithm satisfies Definition 4 since it only applies to data points of non-zero probability. Thus, the algorithm satisfies Definition 4 but not Definition 1. □

We can get a similar definition that is equivalent to differential privacy by looking at all populations $\mathcal{P}$, where the populations determine various joint distributions over data points.

Definition 5 (equivalent)**.** A randomized algorithm $\mathcal{A}$ is said to be *ϵ-differentially private as universal qualified conditioning on the whole database* if for all populations $\mathcal{P}$, if for all databases $d, d' \in \mathcal{D}^n$ at Hamming distance at most 1, and for all output values o, if $\Pr_{\mathcal{P}}[D{=}d] > 0$ and $\Pr_{\mathcal{P}}[D{=}d'] > 0$ then

$$\Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid D{=}d] \leq e^{\epsilon} * \Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid D{=}d'] \tag{13}$$

where $O = \mathcal{A}(D)$ and $D = \langle D_1,\ldots,D_n\rangle$.

Proposition 5. Definitions 1 and 5 are equivalent.

Proof. Definitions 1 implies Definition 5 by the same reasoning as in the proof of Proposition 4.

Assume Definition 5 holds. Let $\mathcal{P}$ be a population that is i.i.d. and assigns non-zero probabilities to all the sequences of n data points. Consider any index i, data points $d_1,\ldots,d_n$ in $\mathcal{D}^n$ and d_i' in $\mathcal{D}$, and output o. $\mathcal{P}$ is such that $\Pr_{\mathcal{P}}[D_1{=}d_1,\ldots,D_n{=}d_n] > 0$ and $\Pr_{\mathcal{P}}[D_1{=}d_1,\ldots,D_i{=}d_i',\ldots,D_n{=}d_n] > 0$ both hold. Thus, since Definition 5 holds for $\mathcal{P}$,

$$\Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid D_1{=}d_1, \ldots, D_n{=}d_n] \leq e^{\epsilon} * \Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid D_1{=}d_1, \ldots, D_i{=}d'_i, \ldots, D_n{=}d_n] \tag{14}$$

$$\Pr_{\mathcal{A}}[\mathcal{A}(d_1, \ldots, d_n){=}o] \leq e^{\epsilon} * \Pr_{\mathcal{A}}[\mathcal{A}(d_1, \ldots, d_i, \ldots, d_n){=}o] \tag{15}$$

where the second line follows from (2). Thus, Definition 1 holds. □

Definition 4 does a reasonable job making precise the intuition behind idea that changing the value of a single D_i in the database D leads to a small change in the distribution on outputs O. As the informal claim is informally an implication of differential privacy, the formal Definition 4 is a formal implication of the differential privacy. Definition 5 shows how to get an equivalence out of a similar definition. However, both of these definitions require conditioning upon the whole database, which seems to be a bit much for discussing the change to a single data point.

5. Differential Privacy as Association with a Single Data Point

By conditioning upon all the data points, Definitions 4 and 5 do not clearly show that the comparison rests on changing the value of a single database input D_i. Let us consider limiting the conditioning to just the changed value D_i.

Definition 6 (neither implied nor implies)**.** A randomized algorithm $\mathcal{A}$ is said to be ϵ*-differentially private as qualified conditioning on a data point* if for all i, for all data points d_i and d'_i in $\mathcal{D}$, and for all output values o, if $\Pr[D_i{=}d_i] > 0$ and $\Pr[D_i{=}d'_i] > 0$ then

$$\Pr[O{=}o \mid D_i{=}d_i] \leq e^{\epsilon} * \Pr[O{=}o \mid D_i{=}d'_i] \tag{16}$$

where $O = \mathcal{A}(D)$ and $D = \langle D_1, \ldots, D_n \rangle$.

Definition 6 does not imply Definition 1 for the same reason Definition 4 does not imply Definition 1: the behavior of the algorithm $\mathcal{A}$ is unconstrained on data points with zero probability while differential privacy (Definition 1) constrains the behavior of the algorithm for even these data points. However, this definition is even further from differential privacy in that differential privacy also does not imply it, meaning it is not even an accurate depiction of the consequences of differential privacy. The reason Definition 1 does not imply Definition 6 is that conditioning upon $D_i = d_i$ or $D_i = d'_i$ might provide information about other data points.

Proposition 6. Definition 1 does not imply Definition 6, nor the other way around.

Proof. Definition 6 does not imply Definition 1 by the same reasoning as Definition 4 does not imply Definition 1 (the proof for Proposition 4) since that proof already uses a database of only a single data point.

To show that Definition 1 does not imply Definition 6, consider an algorithm $\mathcal{A}$ that has ϵ-differential privacy (Definition 1) from using the Laplace Mechanism with ϵ noise for the sum of inputs (count of non-zero inputs) [10]. Further, consider a population $\mathcal{P}$ that is uniform over binary data points but not i.i.d. over $n > 1$ data points. In particular, suppose that data points have zero probability when they are not all equal. That is, $D_1 = D_2 = \cdots = D_n$ and $\Pr_{\mathcal{P}}[D_i{=}0 \mid D_j{=}0] = 1$ and $\Pr_{\mathcal{P}}[D_i{=}1 \mid D_j{=}1] = 1$ for all i and j. (For some settings this counterexample might be unrealistic, raising the question of whether the implication will continue to not hold if we only allow two data points to be equal. Appendix C shows that it will.)

$$\Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid D_n{=}d_n] = \sum_{\langle d_2,\ldots,d_n\rangle \in \mathcal{D}^{n-1}} \Pr_{\mathcal{P}}[\wedge_{i=2}^{n} D_i{=}d_i \mid D_n{=}d_n] * \Pr_{\mathcal{A}}[\mathcal{A}(d_1, d_2, \ldots, d_n){=}o] \tag{17}$$

$$= 1 * \Pr_{\mathcal{A}}[\mathcal{A}(d_n, d_n, \ldots, d_n){=}o] + \sum_{\langle d_1,d_2,\ldots,d_{n-1}\rangle \in \mathcal{D}^{n-1} \text{ s.t. } \exists i \text{ s.t. } d_i \neq d_n} 0 * \Pr_{\mathcal{A}}[\mathcal{A}(d_1, d_2, \ldots, d_n){=}o] \tag{18}$$

$$= \Pr_{\mathcal{A}}[\mathcal{A}(d_n, d_n, \ldots, d_n){=}o] \tag{19}$$

where (17) follows from Lemma 1 in Appendix B and (18) follows from $\Pr_{\mathcal{P}}[\wedge_{i=2}^{n} D_i{=}d_i \mid D_n{=}d_n]$ being 0 whenever D_i is not d_n for any i. Similarly,

$$\Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid D_n{=}d'_n] = \Pr_{\mathcal{A}}[\mathcal{A}(d'_n, d'_n, \ldots, d'_n){=}o] \tag{20}$$

Since $\mathcal{A}$ is the Laplace Mechanism with ϵ noise, for $d_n = 0$, $d'_n = 1$, and $o = 0$,

$$\Pr_{\mathcal{A}}[\mathcal{A}(d_n, d_n, \ldots, d_n){=}o] = e^{n*\epsilon} * \Pr_{\mathcal{A}}[\mathcal{A}(d'_n, d'_n, \ldots, d'_n){=}o] \tag{21}$$

Since $e^{n*\epsilon} > e^{\epsilon}$, the needed bound does not hold:

$$\Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid D_n{=}d_n] = e^{n*\epsilon} * \Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid D_n{=}d'_n] \tag{22}$$

$$> e^{\epsilon} * \Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid D_n{=}d'_n] \tag{23}$$

□

This second issue of conditioning upon $D_i = d_i$ or $D_i = d'_i$ providing information about other data points does not go away if we qualify over all populations in hopes of making a definition equivalent to differential privacy as we did before.

Definition 7 (too strong)**.** A randomized algorithm $\mathcal{A}$ is said to be *ϵ-differentially private as universal qualified conditioning on a data point* if for all populations $\mathcal{P}$, for all i, for all data points d_i and d'_i in $\mathcal{D}$, and for all output values o, if $\Pr_{\mathcal{P}}[D_i{=}d_i] > 0$ and $\Pr_{\mathcal{P}}[D_i{=}d'_i] > 0$ then

$$\Pr_{\mathcal{P}}[O{=}o \mid D_i{=}d_i] \leq e^{\epsilon} * \Pr_{\mathcal{P}}[O{=}o \mid D_i{=}d'_i] \tag{24}$$

where $O = \mathcal{A}(D)$ and $D = \langle D_1, \ldots, D_n\rangle$.

Rather than being equivalent to Definition 1, Definition 7 is strictly stronger than it, and, thus, neither a good characterization of differential privacy nor its consequences.

Proposition 7. Definition 7 implies Definition 1, but not the other way around.

Proof. Definition 1 does not imply Definition 7 by the same reasoning that Definition 1 does not imply Definition 6 (Proposition 6).

To show that Definition 7 implies Definition 1, assume that $\mathcal{A}$ satisfies Definition 7. Choose any $d_1, \cdots, d_n \in \mathcal{D}^n$ and $d'_i \in \mathcal{D}$. Choose $\mathcal{P}$ such that

$$\Pr_{\mathcal{P}}[D_1{=}d_1, \ldots, D_i{=}d_i, \ldots D_n{=}d_n] = \Pr_{\mathcal{P}}[D_1{=}d_1, \ldots, D_i{=}d'_i, \ldots, D_n{=}d_n] = \tfrac{1}{2} \tag{25}$$

For this distribution, $\Pr_{\mathcal{P}}(D_i = d_i) = \Pr_{\mathcal{P}}(D_i = d'_i) = \frac{1}{2}$, and for any o

$$\begin{aligned}
\Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid D_i{=}d'_i] &= \frac{\Pr_{\mathcal{P},\mathcal{A}}[O{=}o \wedge D_i{=}d'_i]}{\Pr_{\mathcal{P}}[D_i{=}d'_i]} && (26)\\
&= \frac{\frac{1}{2} * \Pr_{\mathcal{A}}[\mathcal{A}(d_1, \ldots, d'_i, \ldots, d_n){=}o]}{\frac{1}{2}} && (27)\\
&= \Pr_{\mathcal{A}}[\mathcal{A}(d_1, \ldots, d'_i, \ldots, d_n){=}o] && (28)
\end{aligned}$$

where (27) comes from the randomization of the algorithm being independent of the population. Similarly,

$$\Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid D_i{=}d_i] = \Pr_{\mathcal{A}}[\mathcal{A}(d_1, \ldots, d_i, \ldots, d_n){=}o] \tag{29}$$

Thus for any o,

$$\begin{aligned}
\Pr_{\mathcal{A}}[\mathcal{A}(d_1, \ldots, d_i, \ldots, d_n){=}o] &= \Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid D_i{=}d_i] && (30)\\
&\leq e^{\epsilon} * \Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid D_i{=}d'_i] && (31)\\
&= e^{\epsilon} * \Pr_{\mathcal{A}}[\mathcal{A}(d_1, \ldots, d'_i, \ldots, d_n){=}o] && (32)
\end{aligned}$$

where (31) holds as $\mathcal{A}$ satisfies Definition 7. Together (30) and (32) show that $\mathcal{A}$ satisfies Definition 1. □

To remove the possibility of conditioning upon $D_i = d_i$ or $D_i = d'_i$ providing information about other data points, we can add a new condition that the data points are independent of one another.

Definition 8 (implied, but weaker)**.** A randomized algorithm $\mathcal{A}$ is said to be *ϵ-differentially private as qualified conditioning on an independent data point* if for the given population $\mathcal{P}$, if the D_i is independent of D_j conditioning upon a subset of other data points for all $i \neq j$, for all i, for all data points d_i and d'_i in $\mathcal{D}$, and for all output values o, if $\Pr_{\mathcal{P}}[D_i{=}d_i] > 0$ and $\Pr_{\mathcal{P}}[D_i{=}d'_i] > 0$ then

$$\Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid D_i{=}d_i] \leq e^{\epsilon} * \Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid D_i{=}d'_i] \tag{33}$$

where $O = \mathcal{A}(D)$ and $D = \langle D_1, \ldots, D_n \rangle$.

Proposition 8. Definition 1 implies Definition 8, but not the other way around.

Proof. Definition 8 does not imply Definition 1 by the same reasoning as Definition 4 does not imply Definition 1 (the proof for Proposition 4) since that proof already uses a database of only a single data point.

Definition 1 implies Definition 8 as follows:

$$\Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid D_i{=}d_i] \tag{34}$$

$$= \sum_{\langle d_1,\ldots,d_{i-1},d_{i+1},\ldots,d_n\rangle \in \mathcal{D}^{n-1}} \Pr_{\mathcal{P}}\left[\wedge_{j\in\{1,\ldots,d_{i-1},d_{i+1},\ldots,n\}} D_j{=}d_j \mid D_i{=}d_i\right] * \Pr_{\mathcal{A}}[\mathcal{A}(d_1,\ldots,d_i,\ldots,d_n){=}o] \tag{35}$$

$$= \sum_{\langle d_1,\ldots,d_{i-1},d_{i+1},\ldots,d_n\rangle \in \mathcal{D}^{n-1}} \Pr_{\mathcal{P}}\left[\wedge_{j\in\{1,\ldots,d_{i-1},d_{i+1},\ldots,n\}} D_j{=}d_j \mid D_i{=}d'_i\right] * \Pr_{\mathcal{A}}[\mathcal{A}(d_1,\ldots,d_i,\ldots,d_n){=}o] \tag{36}$$

$$\leq \sum_{\langle d_1,\ldots,d_{i-1},d_{i+1},\ldots,d_n\rangle \in \mathcal{D}^{n-1}} \Pr_{\mathcal{P}}\left[\wedge_{j\in\{1,\ldots,d_{i-1},d_{i+1},\ldots,n\}} D_j{=}d_j \mid D_i{=}d'_i\right] * e^{\epsilon} * \Pr_{\mathcal{A}}[\mathcal{A}(d_1,\ldots,d'_i,\ldots,d_n){=}o] \tag{37}$$

$$= e^{\epsilon} * \Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid D_i{=}d'_i] \tag{38}$$

where (35) and (38) follow from Lemma 1 in the Appendix B, (36) follows from the assumption of independence of D_i from D_j for $j \neq i$, and (37) follows from $\mathcal{A}$ having differential privacy. □

To get a definition equivalent to differential privacy, we look at all the populations $\mathcal{P}$ where the data points are independent of one another.

Definition 9 (equivalent)**.** A randomized algorithm $\mathcal{A}$ is said to be *ϵ-differentially private as universal qualified conditioning on an independent data point* if for all populations $\mathcal{P}$ where the D_i is independent of D_j conditioning upon subset of other data points for all $i \neq j$, for all i, for all data points d_i and d'_i in $\mathcal{D}$, and for all output values o, if $\Pr_{\mathcal{P}}[D_i{=}d_i] > 0$ and $\Pr_{\mathcal{P}}[D_i{=}d'_i] > 0$ then

$$\Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid D_i{=}d_i] \leq e^{\epsilon} * \Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid D_i{=}d'_i] \tag{39}$$

where $O = \mathcal{A}(D)$ and $D = \langle D_1,\ldots,D_n\rangle$.

Proposition 9. Definitions 1 and 9 are equivalent.

Proof. Definition 9 implies Definition 1 by the same reasoning that Definition 7 implies Definition 1 (Proposition 7).

Definition 1 implies Definition 9 by the same reasoning that Definition 1 implies Definition 8 (Proposition 8). □

We see that even if $\mathcal{A}$ has differential privacy under Definition 1, it might not satisfy Definition 6 since learning that $D_i = d_i$ might shed light on other inputs D_j where $j \neq i$. However, if we rule out that possibility, as in Definition 9, the result holds. This issue corresponds to the claim found in some papers that differential privacy has an implicit assumption of independence between data points [1, 5]. In particular, Proposition 9 is nearly identical to Theorem 6.1 from [1]. A minor difference is that our Definition 9 does not require (39) to hold for points with zero probability, as the probabilities are undefined for such

points. We believe this condition to have been implicitly assumed in their work as well.

We will show a way of removing the limitation to independent data points by viewing differential privacy as causal property. Thus, rather than interpret this limitation as an implicit assumption of differential privacy, we view it as indicative of how differential privacy is rather better understood as a causal property than as an property about association or independence.

6. Differential Privacy as Causation on the Whole Database

Due to differential privacy's behavior on associated inputs and its requirement of considering zero-probability database values, differential privacy is not a straightforward property about the independence or degree of association of the database and the algorithm's output. The would-be conditioning upon zero-probability values corresponds to a form of counterfactual reasoning asking what the algorithm would had performed had the database taken on a particular value that it might never actually take on. Experiments with such counterfactuals that may never naturally occur form the core of causation. The behavior of differential privacy on associated inputs corresponds to the atomicity assumption found in causal reasoning, that one can change the value of an input without changing the values of other inputs. (More generally, atomicity, implicit in the structural equation approach to defining causation, allows one to ask what would happen if the value of a variable changed independently of changes to any other variables that are not affected by the changed variable.) With these motivations, we will show that differential privacy is equivalent to a causal property that makes the change in a single data point explicit.

Before doing so, we will introduce a framework for precisely reasoning about causation based upon Pearl's [11] and show an equivalence between differential privacy and a causal property on the whole database to echo Proposition 5. The causal equivalence here is simpler than that with Definition 5 since it does not need qualifications around zero probability data points, which removes the need to quantify over all populations.

To develop such a causal interpretation of differential privacy, we start by re-interpreting the equation $O = \mathcal{A}(D)$. Previously, we viewed it as shorthand for an observation that two random variables O and $\mathcal{A}(D)$ are related such that $O(\omega) = \mathcal{A}(D(\omega))(\omega)$, which says nothing about why this relation holds. Now, we interpret it as a stronger causal relation asserting that the value of the output O is caused by the value of the input D, that is, as a causal structural equation. We will denote this interpretation by $O := \mathcal{A}(D)$ since it is closer to an assignment than equality due to its directionality. In particular, the value of O might change if the value of D is artificially altered (e.g., by random assignment in an experiment) but the value of D would not change if O is artificially altered since causation only flows from causes to effects. To make this more precise, let $\mathrm{do}(D{=}d)$ denote an intervention setting the value of D to d

(Pearl's *do* notation [11]). Using this notation, $\Pr[O{=}o \mid \mathrm{do}(D{=}d)]$ represents what the probability of $O = o$ would be if the value of D were set to d by intervention. Similar to normal conditioning on $D = d$, $\Pr[O{=}o \mid \mathrm{do}(D{=}d)]$ need not equal $\Pr[O{=}o]$. However, $\Pr[D{=}d \mid \mathrm{do}(O{=}o)] = \Pr[D{=}d]$ since O is downstream of D, and, thus, changing O would have no affects on D.

Similarly, we replace $D = \langle D_1, \ldots, D_n \rangle$ with $D := \langle D_1, \ldots, D_n \rangle$. That is, we consider the value of the whole database to be caused by the values of its data points and nothing more. Furthermore, we require that the $D_1, \ldots, D_n$ only cause D and does not have any other effect. In particular, we do not allow D_i to affect D_j for $i \neq j$. This requirement might seem to prevent one person's attribute from affecting another's, for example, prevent one person's race from affecting his child's race. This is not case since $D_1, \ldots, D_n$ represent the data points provided as inputs to the algorithm and not the actual attributes themselves. One could model these attributes, such as race itself, as random variables $R_1, \ldots, R_n$ where $D_i := R_i$ for all i and allow R_i to affect R_j without changing our results. For example, the following causal diagram is acceptable: However, since we are not focusing on the causes of $D_1, \ldots, D_n$,

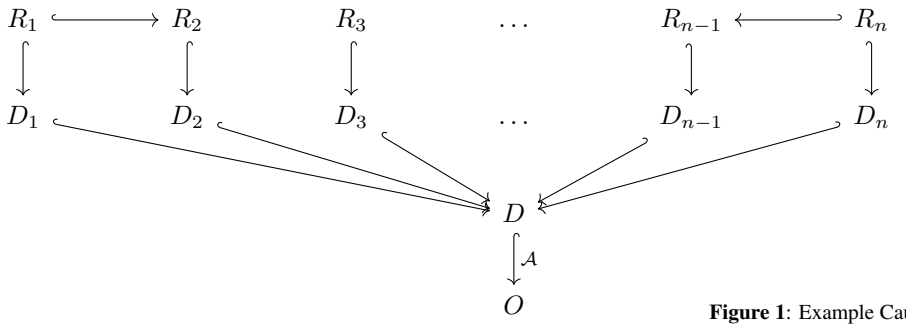

Figure 1: Example Causal Diagram. The arrows $\hookrightarrow$ represent causal relations. The variable at the start of the arrow affects the variable at the end of the arrow. For example, R_2 is caused by R_1. The absence of an arrow from one variable to another means the first does affect the second.

we will model using a probability distribution over their values. Reflecting that they might that their causes (e.g., $R_1, \ldots, R_n$) might have causal relations, we do not require the distributions over $D_1, \ldots, D_n$ to be independent.

Recall that $\Pr[O{=}o]$ is the probability of the algorithm's output being o under the naturally occurring distribution of inputs (and coin flips internal to $\mathcal{A}$), that $\Pr[O{=}o \mid D_i{=}d_i]$ is that probability conditioned upon seeing $D_i = d_i$, and that $\Pr[O{=}o \mid \mathrm{do}(D_i{=}d_i)]$ represents the probability of $O = o$ given an intervention setting the value of D_i to d_i. The last probability depends upon how the intervention on D_i will flow downstream to D and then O. The probability differs from the conditional probability in that setting D_i to d_i provides no information about D_j for $j \neq i$ whereas if D_i and D_j are associated, then seeing the value D_i does provide information about D_j. Intuitively, this lack of information is because the artificial setting of D_i to d_i has no causal influence on D_j due to the data points not affecting one another and the artificial setting, by being artificial, tells us nothing about the associations found in the naturally occurring world. On the other hand, artificially setting R_1 to r_1 in the causal diagram above (fig. 1) will provide information about D_2

since R_1 has an affect on D_2 in addition to D_1. A second difference is that $\Pr[O{=}o \mid \mathrm{do}(D_i{=}d_i)]$ is defined even when $\Pr[D_i{=}d_i]$ is zero. Importantly, interventions on D_is may not accurately model the choice an individual has to make while providing their attributes, or any other realizable mechanism for modifying their attributes. Instead, interventions on D_i model changing the values provided as input to the algorithm which are naturally change-able without affecting other values in the world.

With the machinery in place to reason about causation, we can get a definition equivalent to differential privacy very easily.

Definition 10 (equivalent)**.** A randomized algorithm $\mathcal{A}$ is said to be *ϵ-differentially private as intervention on the whole database* if for all i, for all data points $d_1, \ldots, d_n$ in $\mathcal{D}^n$ and d_i' in $\mathcal{D}$, and for all output values o,

$$\Pr[O{=}o \mid \mathrm{do}(D_1{=}d_1, \ldots, D_n{=}d_n)] \leq e^{\epsilon} * \Pr[O{=}o \mid \mathrm{do}(D_1{=}d_1, \ldots, D_i{=}d_i', \ldots, D_n{=}d_n)] \tag{40}$$

where $O := \mathcal{A}(D)$ and $D := \langle D_1, \ldots, D_n \rangle$.

Proposition 10. Definitions 1 and 10 are equivalent.

Proof. $\Pr[O{=}o \mid \mathrm{do}(D_1{=}d_1, \ldots, D_n{=}d_n)] = \Pr[\mathcal{A}(d_1, \ldots, d_n){=}o]$ and

$\Pr[O{=}o \mid \mathrm{do}(D_1{=}d_1, \ldots, D_i{=}d_i', \ldots, D_n{=}d_n)] = \Pr[\mathcal{A}(d_1, \ldots, d_i, \ldots, d_n){=}o]$

from Lemma 2 in Appendix D. □

The simple Definition 10 works whereas our attempts with conditional probabilities require considerable complexity because we can causally fix data points to values with zero probability. For completeness, we will state a more complex definition that quantifies over all populations:

Definition 11 (equivalent)**.** A randomized algorithm $\mathcal{A}$ is said to be *ϵ-differentially private as universal intervention on the whole database* if for all populations $\mathcal{P}$, for all i, for all data points $d_1, \ldots, d_n$ in $\mathcal{D}^n$ and d_i' in $\mathcal{D}$, and for all output values o,

$$\Pr[O{=}o \mid \mathrm{do}(D_1{=}d_1, \ldots, D_n{=}d_n)] \leq e^{\epsilon} * \Pr[O{=}o \mid \mathrm{do}(D_1{=}d_1, \ldots, D_i{=}d_i', \ldots, D_n{=}d_n)] \tag{41}$$

where $O := \mathcal{A}(D)$ and $D := \langle D_1, \ldots, D_n \rangle$.

Proposition 11. Definitions 1 and 11 are equivalent.

Proof. The proof follows in the same manner as Proposition 10 since that proof applies to all populations $\mathcal{P}$. □

However, Definitions 10 and 11, by fixing every data point, do not capture the local nature of the decision facing a single potential survey participant.

7. Differential Privacy as Causation on a Single Data Point

We can define a notion similar to differential privacy that uses a causal intervention on a single data point as follows:

Definition 12 (implied, but weaker)**.** Given a population $\mathcal{P}$, a randomized algorithm $\mathcal{A}$ is said to be *ϵ-differentially private as intervention on a data point* if for all i, for all data points d_i and d'_i in $\mathcal{D}$, and for all output values o,

$$\Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid \mathrm{do}(D_i{=}d_i)] \leq e^{\epsilon} * \Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid \mathrm{do}(D_i{=}d'_i)] \tag{42}$$

where $O := \mathcal{A}(D)$ and $D := \langle D_1, \ldots, D_n \rangle$.

This definition is implied by differential privacy, but it does not imply differential privacy. The reason is similar to why Definitions 4 and 6 do not imply differential privacy (Propositions 4 and 6) in that they all involve a counterexample with a population $\mathcal{P}$ that hides the effects of a possible value of the data point by assigning the value a probability of zero. For the associative definition, the counterexample involves only a single data point, but, for this causal definition, the counterexample has to have two data points. The reason is that, since the do operation acts on a single data point at a time, it can flush out the effects of a single zero-probability value but not the interactions between two zero-probability values.

Proposition 12. Definition 1 implies Definition 12, but not the other way around.

Proof. W.l.o.g., assume $i = n$.

Assume Definition 1 holds. Then,

$$\Pr[\mathcal{A}(d_1, \ldots, d_{n-1}, d_n){=}o] \leq e^{\epsilon} * \Pr[\mathcal{A}(d_1, \ldots, d_{n-1}, d'_n){=}o] \tag{43}$$

for all $d_1, \ldots, d_n$ in $\mathcal{D}^n$ and d'_n in $\mathcal{D}$. This implies that for any $\mathcal{P}$,

$$\Pr_{\mathcal{P}}\left[\wedge_{i=1}^{n-1} D_i{=}d_i\right] * \Pr_{\mathcal{A}}[\mathcal{A}(d_1, \ldots, d_{n-1}, d_n){=}o] \leq e^{\epsilon} * \Pr_{\mathcal{P}}\left[\wedge_{i=1}^{n-1} D_i{=}d_i\right] * \Pr_{\mathcal{A}}[\mathcal{A}(d_1, \ldots, d_{n-1}, d'_n{=}o] \tag{44}$$

for all $d_1, \ldots, d_n$ in $\mathcal{D}^n$ and d'_n in $\mathcal{D}$. Thus,

$$\sum_{\langle d_1,\ldots,d_{n-1}\rangle \in \mathcal{D}^{n-1}} \Pr_{\mathcal{P}}\left[\wedge_{i=1}^{n-1} D_i{=}d_i\right] * \Pr[\mathcal{A}(d_1, \ldots, d_{n-1}, d_n){=}o] \leq \sum_{\langle d_1,\ldots,d_{n-1}\rangle \in \mathcal{D}^{n-1}} e^{\epsilon} * \Pr_{\mathcal{P}}\left[\wedge_{i=1}^{n-1} D_i{=}d_i\right] * \Pr[\mathcal{A}(d_1, \ldots, d_{n-1}, d'_n){=}o] \tag{45}$$

$$\sum_{\langle d_1,\ldots,d_{n-1}\rangle \in \mathcal{D}^{n-1}} \Pr_{\mathcal{P}}\left[\wedge_{i=1}^{n-1} D_i{=}d_i\right] * \Pr[\mathcal{A}(d_1, \ldots, d_{n-1}, d_n){=}o] \leq e^{\epsilon} * \sum_{\langle d_1,\ldots,d_{n-1}\rangle \in \mathcal{D}^{n-1}} \Pr_{\mathcal{P}}\left[\wedge_{i=1}^{n-1} D_i{=}d_i\right] * \Pr[\mathcal{A}(d_1, \ldots, d_{n-1}, d'_n){=}o] \tag{46}$$

$$\Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid \mathrm{do}(D_n{=}d_n)] \leq e^{\epsilon} * \Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid \mathrm{do}(D_n{=}d'_n)] \tag{47}$$

where the last line follows from Lemma 3 in Appendix D.

Definition 12 is, however, weaker than differential privacy. Consider the case of a database holding two data points whose value could be 0, 1, or 2. Suppose the population $\mathcal{P}$ is such that $\Pr[D_1{=}2] = 0$ and $\Pr[D_2{=}2] = 0$. Consider an algorithm $\mathcal{A}$ such that

$$\Pr_{\mathcal{A}}[\mathcal{A}(d_1,d_2){=}0] = 1/2 \qquad \Pr_{\mathcal{A}}[\mathcal{A}(d_1,d_2){=}1] = 1/2 \qquad \text{when } d_1 \neq 2 \text{ or } d_2 \neq 2 \tag{48}$$

$$\Pr_{\mathcal{A}}[\mathcal{A}(2,2){=}0] = 1 \qquad \Pr_{\mathcal{A}}[\mathcal{A}(2,2){=}1] = 0 \tag{49}$$

The algorithm does not satisfy Definition 1 due to its behavior when both of the inputs are 2.

However, using Lemma 3 in Appendix D,

$$\Pr_{\mathcal{P},\mathcal{A}}[O{=}0 \mid \mathrm{do}(D_1{=}0)] = 1/2 \quad \Pr_{\mathcal{P},\mathcal{A}}[O{=}1 \mid \mathrm{do}(D_1{=}0)] = 1/2 \tag{50}$$

$$\Pr_{\mathcal{P},\mathcal{A}}[O{=}0 \mid \mathrm{do}(D_1{=}1)] = 1/2 \quad \Pr_{\mathcal{P},\mathcal{A}}[O{=}1 \mid \mathrm{do}(D_1{=}1)] = 1/2 \tag{51}$$

$$\Pr_{\mathcal{P},\mathcal{A}}[O{=}0 \mid \mathrm{do}(D_1{=}2)] = 1/2 \quad \Pr_{\mathcal{P},\mathcal{A}}[O{=}1 \mid \mathrm{do}(D_1{=}2)] = 1/2 \tag{52}$$

since $\Pr_{\mathcal{P}}[D_2{=}2] = 0$. A similar result holds switching the roles of D_1 and D_2. Thus, the algorithm satisfies Definition 12 for $\mathcal{P}$ but not Definition 1.

□

Despite being only implied by, not equivalent to, differential privacy, Definition 12 captures the intuition behind the sentence

> This definition states that changing the value of a single D_i in the database D leads to a small change in the distribution on outputs O.

when viewed as characterizing the implications of differential privacy. To get an equivalence, we can quantify over all populations as we did to get an equivalence for association, but this time we need not worry about zero-probability data points or independence. This simplifies the definition and makes it a more natural characterization of differential privacy.

Definition 13 (equivalent)**.** A randomized algorithm $\mathcal{A}$ is said to be *ϵ-differentially private as universal intervention on a data point* if all populations $\mathcal{P}$, for all i, for all data points d_i and d_i' in $\mathcal{D}$, and for all output values o,

$$\Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid \mathrm{do}(D_i{=}d_i)] \leq e^{\epsilon} \Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid \mathrm{do}(D_i{=}d_i')] \tag{53}$$

where $O := \mathcal{A}(D)$ and $D := \langle D_1, \ldots, D_n \rangle$.

Proposition 13. Definitions 1 and 13 are equivalent.

Proof. W.l.o.g. and simplicity in notation, assume $i = n$.

Assume Definition 1 holds. The needed result follows from Proposition 12.

Assume Definition 13 holds. Then, for all $\mathcal{P}$,

$$\Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid \mathrm{do}(D_i{=}d_i)] \leq e^{\epsilon} * \Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid \mathrm{do}(D_i{=}d_i')] \tag{54}$$

$$\sum_{\langle d_1,\ldots,d_{n-1}\rangle \in \mathcal{D}^{n-1}} \Pr_{\mathcal{P}}\left[\wedge_{i=1}^{n-1} D_i{=}d_i\right] * \Pr[\mathcal{A}(d_1,\ldots,d_{n-1},d_n){=}o] \leq e^{\epsilon} * \sum_{\langle d_1,\ldots,d_{n-1}\rangle \in \mathcal{D}^{n-1}} \Pr_{\mathcal{P}}\left[\wedge_{i=1}^{n-1} D_i{=}d_i\right] * \Pr[\mathcal{A}(d_1,\ldots,d_{n-1},d_n'){=}o] \tag{55}$$

follows from Lemma 3 in Appendix D.

For any $d_1^\dagger, \ldots, d_{n-1}^\dagger$ in $\mathcal{D}^{n-1}$, let $\mathcal{P}^{d_1^\dagger,\ldots,d_{n-1}^\dagger}$ be such that

$$\Pr_{\mathcal{P}^{d_1^\dagger,\ldots,d_{n-1}^\dagger}}\left[\wedge_{i=1}^{n-1} D_i{=}d_i^\dagger\right] = 1 \tag{56}$$

For any $d_1^\dagger, \ldots, d_n^\dagger$ in $\mathcal{D}^n$ and d'_n in $\mathcal{D}$, (55) implies

$$\begin{aligned}&\sum_{\langle d_1,\ldots,d_{n-1}\rangle \in \mathcal{D}^{n-1}} \Pr_{\mathcal{P}^{d_1^\dagger,\ldots,d_{n-1}^\dagger}}\left[\wedge_{i=1}^{n-1} D_i{=}d_i\right] * \Pr_{\mathcal{A}}[\mathcal{A}(d_1,\ldots,d_{n-1},d_n^\dagger){=}o] \\ &\quad \leq e^\epsilon \sum_{\langle d_1,\ldots,d_{n-1}\rangle \in \mathcal{D}^{n-1}} \Pr_{\mathcal{P}^{d_1^\dagger,\ldots,d_{n-1}^\dagger}}\left[\wedge_{i=1}^{n-1} D_i{=}d_i\right] * \Pr_{\mathcal{A}}[\mathcal{A}(d_1,\ldots,d_{n-1},d'_n){=}o]\end{aligned} \tag{57}$$

Thus,

$$\Pr_{\mathcal{A}}[\mathcal{A}(d_1^\dagger,\ldots,d_{n-1}^\dagger,d_n^\dagger){=}o] \leq e^\epsilon \Pr_{\mathcal{A}}[\mathcal{A}(d_1^\dagger,\ldots,d_{n-1}^\dagger,d'_n){=}o] \tag{58}$$

since both sides has a non-zero probability for $\Pr_{\mathcal{P}^{d_1^\dagger,\ldots,d_{n-1}^\dagger}}\left[\wedge_{i=1}^{n-1} D_i{=}d_i\right]$ at just the single sequences of data point values $d_1^\dagger, \ldots, d_{n-1}^\dagger$. □

8. Conclusion and Discussion

We have shown that is possible to view differential privacy as an associative property with an independence assumption but that it is cleaner to view it as a causal property. We believe this helps to explain why some researchers feel that differential privacy requires an assumption of independence while other researchers do not.

Our observation also reduces the benefits and drawbacks of each camp's view to those known from studying association and causation. For example, the first camp's causal view only requires looking at the system itself (causation is an inherent property of systems) This difference explains why the second camp speaks of the distribution over data points despite the definition of differential privacy not mentioning it.

The causal characterization also requires us to distinguish between an individual's attributes (R_is) and the data that is input to an algorithm (D_is), and intervenes on the latter. Under the assumption that individuals don't lie, the associative interpretation does not require this distinction since conditioning on one is identical to conditioning the other. This distinction captures an aspect of the difference between protecting "secrets about you" (R_i) and protecting "secrets from you" (D_i) pointed out by the first camp [8, 9], where differential privacy protects the latter in a causal sense.

We believe these results have implications beyond explaining the differences between these two camps. Having shown a precise sense in which differential privacy is a causal property, we can use all the results of statistics, experimental design, and science about causation while studying differential

privacy. For example, Tang et al. studies Apple's claim that MacOS uses differential privacy and attempt to reverse engineer the degree ϵ of privacy used by Apple from the compiled code and configuration files [14]. Consider a version of this problem in which the system purportedly providing differential privacy is a server controlled by some other entity. In this case, the absence of code and configuration files necessities a blackbox investigation of the system. From the outside, we can study whether such a system has differential privacy as advertised by using experiments and significance testing [15] similar to how Tschantz et al.'s prior work uses it for studying information flow [16]. (For an application, see [17].) Alternately, using the associative view, we could approach the problem using observational studies.

In the opposite direction, the natural sciences can use differential privacy as an effect-size metric, which would inherit all the pleasing properties known of differential privacy. For example, differential privacy composes cleanly with itself, both in sequence and in parallel [18]. The same results would also apply to the effect-size metric that differential privacy suggests.

Acknowledgements. We thank Deepak Garg for conversations about causation and Arthur Azevedo de Amorim for comments on a draft. We gratefully acknowledge funding support from the National Science Foundation (Grants 1514509, 1704845, and 1704985) and DARPA (FA8750-16-2-0287). The opinions in this paper are those of the authors and do not necessarily reflect the opinions of any funding sponsor or the United States Government.

Appendices

A. Two Views of Differential Privacy: A Brief History

Here, we briefly recount the history of the two camps surrounding differential privacy. Having not participated in differential privacy's formative years, we welcome refinements to our account.

In 1965, S. L. Warner presented the *randomized response* method of providing differential privacy [19]. In 1977, T. Dalenius presented a different view of privacy, *Semantic Privacy* [20]. The randomized response model and semantic privacy can be viewed as the prototypes of the first and second camps respectively, although these early works appeared to have had little impact on the actual formation of the camps over a quarter century later.

In March 2006, Dwork, McSherry, Nissim, and Smith presented a paper containing the first modern instance of differential privacy under the name of "ϵ-indistinguishable" [10]. The earliest use of the term "differential privacy" comes from an paper by Dwork presented in July 2006 [12]. This paper of Dwork explicitly rejects the second camp (page 8):

> Note that a bad disclosure can still occur [despite differential privacy], but [differential privacy] assures the individual that it will not be the presence of her data

> that causes it, nor could the disclosure be avoided through any action or inaction on the part of the user.

and further contains a proof that Dalenius's Semantic Privacy is impossible. (The proof was joint work with Naor, with whom Dwork later further developed the impossible result [21].) Furthermore, the paper promotes the first camp's view (page 9):

> A mechanism K satisfying [differential privacy] addresses concerns that any participant might have about the leakage of her personal information x: even if the participant removed her data from the data set, no outputs (and thus consequences of outputs) would become significantly more or less likely. For example, if the database were to be consulted by an insurance provider before deciding whether or not to insure Terry Gross, then the presence or absence of Terry Gross in the database will not significantly affect her chance of receiving coverage.

Later works further expound upon their position [22, 23].

In 2011, papers started to question whether differential privacy actually provides a meaningful notion of privacy [24, 1, 25]. These papers point to the fact that a released statistic can enable inferring sensitive information about a person, similar to the attacks Dalenius wanted to prevent [20], even when that statistic was computed using a differentially private algorithm. While the earlier work on differential privacy acknowledged this limitation, these papers provide examples where correlations, or more generally associations, between data points can enable inferences that some people might not expect to be possible under differential privacy. They and later work (e.g., [2, 3, 4, 5]) attempt to find stronger definitions that account for such correlations and provide protections against such inferential threats. In some cases, these authors assert that such inferential threats are violations of privacy and not what people expect of differential privacy. For example, Liu et al.'s abstract states that associations between data points can lead to "degradation in expected privacy levels" [5].

Those promoting the original view of differential privacy have re-asserted that differential privacy was never intended to prevent all inferential privacy threats and that doing so is impossible [6, 7, 8, 9]. McSherry goes the furthest, asserting that inferential privacy is neither privacy nor an appealing concept [8]. He calls it "forgetablity" invoking the European Union's right to be forgotten and points out that preventing inferences prevents people using data and scientific progress. He asserts that people should only have an expectation to the privacy of data they own, not data about them, and that differential privacy captures this concept.

We know of no works from the second camp that have explicitly responded to the first camp's critique of their goals. Thus, we presume that second camp continues to desire a stronger property than differential privacy. We will explore the relationship between the properties desired by each camp in detail below.

B. Calculations for Association

The following lemma aids reasoning about conditioning.

Lemma 1. Let O be the random variable for the output of $\mathcal{A}(D_1, \ldots, D_n)$, then for all o and $d_1, \ldots, d_n$, if $\Pr[D_j{=}d_j] > 0$, then

$$\Pr[O{=}o \mid D_j{=}d_j] = \sum_{\langle d_1,\ldots,d_{j-1},d_{j+1},\ldots,d_n\rangle \in \mathcal{D}^{n-1}} \Pr\left[\wedge_{i\in\{1,\ldots,d_{j-1},d_{j+1},\ldots,n\}} D_i{=}d_i \mid D_j{=}d_j\right] * \Pr[\mathcal{A}(d_1,\ldots,d_n){=}o] \tag{59}$$

Proof. With out loss of generality, we assume $j = n$. Similarly to the above case, for d'_n such that $\Pr[D_n{=}d'_n] > 0$,

$$\begin{aligned}
&\Pr[O{=}o \mid D_n{=}d'_n] && (60)\\
&= \Pr\left[\textstyle\bigvee_{d_1\in\mathcal{D},\ldots,d_n\in\mathcal{D}} O{=}o \wedge \bigwedge_{i=1}^{n} D_i{=}d_i \;\middle|\; D_n{=}d'_n\right] && (61)\\
&= \sum_{\langle d_1,\ldots,d_n\rangle\in\mathcal{D}^n} \Pr\left[O{=}o \wedge \textstyle\bigwedge_{i=1}^{n} D_i{=}d_i \;\middle|\; D_n{=}d'_n\right] && (62)\\
&= \sum_{\langle d_1,\ldots,d_n\rangle\in\mathrm{supp}(\mathcal{D}^n)} \Pr\left[O{=}o \wedge \textstyle\bigwedge_{i=1}^{n} D_i{=}d_i \;\middle|\; D_n{=}d'_n\right] && (63)\\
&= \sum_{\langle d_1,\ldots,d_n\rangle\in\mathrm{supp}(\mathcal{D}^n)} \Pr\left[O{=}o \wedge D_n{=}d'_n \wedge \textstyle\bigwedge_{i=1}^{n} D_i{=}d_i\right] / \Pr[D_n{=}d'_n] && (64)\\
&= \sum_{\langle d_1,\ldots,d_n\rangle\in\mathrm{supp}(\mathcal{D}^n,d'_n)} \Pr\left[O{=}o \wedge D_n{=}d'_n \wedge \textstyle\bigwedge_{i=1}^{n} D_i{=}d_i\right] / \Pr[D_n{=}d'_n] && (65)\\
&= \sum_{\langle d_1,\ldots,d_n\rangle\in\mathrm{supp}(\mathcal{D}^n,d'_n)} \Pr\left[O{=}o \wedge \textstyle\bigwedge_{i=1}^{n} D_i{=}d_i \;\middle|\; D_n{=}d'_n\right] && (66)\\
&= \sum_{\langle d_1,\ldots,d_n\rangle\in\mathrm{supp}(\mathcal{D}^n,d'_n)} \Pr\left[O{=}o \;\middle|\; \textstyle\bigwedge_{i=1}^{n} D_i{=}d_i, D_n{=}d'_n\right] * \Pr\left[\textstyle\bigwedge_{i=1}^{n} D_i{=}d_i \;\middle|\; D_n{=}d'_n\right] && (67)\\
&= \sum_{\langle d_1,\ldots,d_n\rangle\in\mathrm{supp}(\mathcal{D}^n,d'_n)} \Pr\left[O{=}o \;\middle|\; \textstyle\bigwedge_{i=1}^{n-1} D_i{=}d_i, D_n{=}d'_n\right] * \Pr\left[\textstyle\bigwedge_{i=1}^{n-1} D_i{=}d_i \;\middle|\; D_n{=}d'_n\right] && (68)\\
&= \sum_{\langle d_1,\ldots,d_n\rangle\in\mathrm{supp}(\mathcal{D}^n,d'_n)} \Pr[\mathcal{A}(D){=}o] * \Pr\left[\textstyle\bigwedge_{i=1}^{n-1} D_i{=}d_i \;\middle|\; D_n{=}d'_n\right] && (69)\\
&= \sum_{\langle d_1,\ldots,d_n\rangle\in\mathcal{D}^n} \Pr[\mathcal{A}(D){=}o] * \Pr\left[\textstyle\bigwedge_{i=1}^{n-1} D_i{=}d_i \;\middle|\; D_n{=}d'_n\right] && (70)
\end{aligned}$$

where $\mathrm{supp}(\mathcal{D}^n, d'_n)$ is equal to $\{\langle d_1, \ldots, d_n\rangle \in \mathcal{D}^n \mid \Pr[\bigwedge_{i=1}^{n} D_i{=}d_i] > 0 \wedge d_n = d'_n\}$.

□

C. Proof with More Realistic Counterexample

The counterexample used by the proof of Proposition 6 showing that Definition 1 does not imply Definition 6 might appear unrealistic to some readers. Here we redo the proof with a less extreme counterexample but more complex calculations.

Proof. Consider $\mathcal{P}$ that is uniform over binary data points but not i.i.d. In particular, while $D_3, \ldots, D_n$ are i.i.d. and independent of D_1 and D_2, $D_1 = D_2$, that is, $\Pr[D_1{=}0 \mid D_2{=}0] = 1$ and $\Pr[D_1{=}1 \mid D_2{=}1] = 1$.

$$\Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid D_1{=}d_1] \tag{71}$$

$$= \sum_{\langle d_2,\ldots,d_n\rangle \in \mathcal{D}^{n-1}} \Pr_{\mathcal{P}}\left[\wedge_{i=2}^{n} D_i{=}d_i \mid D_1{=}d_1\right] * \Pr[\mathcal{A}(d_1, d_2, d_3, d_4, \ldots, d_n){=}o] \tag{72}$$

$$= \sum_{\langle d_2,\ldots,d_n\rangle \in \mathcal{D}^{n-1}} \Pr_{\mathcal{P}}\left[\wedge_{i=3}^{n} D_i{=}d_i \mid D_1{=}d_1, D_2{=}d_2\right] * \Pr_{\mathcal{P}}[D_2{=}d_2 \mid D_1{=}d_1] * \Pr[\mathcal{A}(d_1, d_2, d_3, d_4, \ldots, d_n){=}o] \tag{73}$$

$$= \sum_{\langle d_3,\ldots,d_n\rangle \in \mathcal{D}^{n-1}} \Pr_{\mathcal{P}}\left[\wedge_{i=3}^{n} D_i{=}d_i \mid D_1{=}d_1, D_2{=}d_1\right] * \Pr[\mathcal{A}(d_1, d_1, d_3, d_4, \ldots, d_n){=}o] \tag{74}$$

$$= \sum_{\langle d_3,\ldots,d_n\rangle \in \mathcal{D}^{n-1}} \Pr_{\mathcal{P}}\left[\wedge_{i=3}^{n} D_i{=}d_i\right] * \Pr[\mathcal{A}(d_1, d_1, d_3, d_4, \ldots, d_n){=}o] \tag{75}$$

$$= \sum_{\langle d_3,\ldots,d_n\rangle \in \mathcal{D}^{n-1}} 1/2^{n-2} * \Pr[\mathcal{A}(d_1, d_1, d_3, d_4, \ldots, d_n){=}o] \tag{76}$$

$$= 1/2^{n-2} * \sum_{\langle d_3,\ldots,d_n\rangle \in \mathcal{D}^{n-1}} \Pr[\mathcal{A}(d_1, d_1, d_3, d_4, \ldots, d_n){=}o] \tag{77}$$

where (75) follows from $D_3, \ldots, D_n$ being independent of D_1 and D_2 and (76) follows from $\mathcal{P}$ being uniform. Similarly,

$$\Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid D_1{=}d_1'] = \sum_{\langle d_3,\ldots,d_n\rangle \in \mathcal{D}^{n-1}} \Pr\left[\wedge_{i=3}^{n} D_i{=}d_i\right] * \Pr[\mathcal{A}(d_1', d_1', d_3, d_4, \ldots, d_n){=}o] \tag{78}$$

$$= 1/2^{n-2} * \sum_{\langle d_3,\ldots,d_n\rangle \in \mathcal{D}^{n-1}} \Pr[\mathcal{A}(d_1', d_1', d_3, d_4, \ldots, d_n){=}o] \tag{79}$$

If $\mathcal{A}$ has ϵ-differential privacy (Definition 1) from using the Laplace Mechanism with ϵ noise for the sum of inputs (count of non-zero inputs) [10], $d_1 = 0$, $d_1' = 1$, and $o = 0$, then

$$\Pr[\mathcal{A}(d_1, d_1, d_3, d_4, \ldots, d_n){=}o] = e^{2\epsilon} * \Pr[\mathcal{A}(d_1', d_1', d_3, d_4, \ldots, d_n){=}o] \tag{80}$$

for all $d_1, \ldots, d_n$ and d_1'.

$$\Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid D_1{=}d_1] = 1/2^{n-2} * \sum_{\langle d_3,\ldots,d_n\rangle \in \mathcal{D}^{n-1}} \Pr[\mathcal{A}(d_1, d_1, d_3, d_4, \ldots, d_n){=}o] \tag{81}$$

$$= 1/2^{n-2} * \sum_{\langle d_3,\ldots,d_n\rangle \in \mathcal{D}^{n-1}} e^{2\epsilon} * \Pr[\mathcal{A}(d_1', d_1', d_3, d_4, \ldots, d_n){=}o] \tag{82}$$

$$= e^{2\epsilon} * 1/2^{n-2} * \sum_{\langle d_3,\ldots,d_n\rangle \in \mathcal{D}^{n-1}} \Pr[\mathcal{A}(d_1', d_1', d_3, d_4, \ldots, d_n){=}o] \tag{83}$$

$$= e^{2\epsilon} * \Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid D_1{=}d_1'] \tag{84}$$

Since $e^{2\epsilon} > e^{\epsilon}$, the needed bound does not hold:

$$\Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid D_1{=}d_1] = e^{2\epsilon} * \Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid D_1{=}d_1'] \tag{85}$$

$$> e^{\epsilon} * \Pr_{\mathcal{P},\mathcal{A}}[O{=}o \mid D_1{=}d_1'] \tag{86}$$

□

D. Details of Causation

To make the above intuitions about causation formal, we use a slight modification of Pearl's causal models.[1] Pearl uses *structural equation models* (SEMs). Let a SEM $M = \langle \mathcal{V}_{\mathsf{en}}, \mathcal{V}_{\mathsf{ex}}, \mathcal{E} \rangle$ includes a set of *variables* partitioned into *endogenous* (or dependent) variables $\mathcal{V}_{\mathsf{en}}$ and *background* (or exogenous, or independent) variables $\mathcal{V}_{\mathsf{ex}}$. M also includes a set $\mathcal{E}$ of structural equations. Each endogenous variable X has a structural equation $X := F_x(\vec{Y})$ where $\vec{Y}$ is a list of other variables other than X and F_X is a possibly randomized function. We call the variables $\vec{Y}$ the *parents* of X and denote them by $\mathsf{pa}(X)$. We call a variable Z an *ancestor* of X if its in the transitive closure of the parents relation with X.

[1] The models we use are suggested by Pearl for handling "inherent" randomness [11, p. 220] and differs from the model he typically uses (his Definition 7.1.6) by allowing randomization in the structural equations F_V. We find this randomization helpful for modeling the randomization with the algorithm $\mathcal{A}$.

As an example of an SEM, consider the M^{dp} that models the setting of differential privacy:

- the background variables $\mathcal{V}_{\mathsf{ex}}^{\mathsf{dp}}$ are $R_1, \ldots, R_n$;
- the endogenous variables $\mathcal{V}_{\mathsf{en}}^{\mathsf{dp}}$ are $D_1, \ldots, D_n$, D, and O;
- the structural equations $\mathcal{E}^{\mathsf{dp}}$ are $D_i := R_i$ for all i, $D := \langle D_1, \ldots, D_n \rangle$, and $O := \mathcal{A}(D)$.

We limit ourselves to *non-recursive* SEMs, those in which the variables may be ordered such that all the background variables come before all the endogenous variables and no variable has a parent that comes before it in the ordering. We may view such SEMs as similar to a program where the background variables are inputs to the program and the ordering determines the order of assignment statements in the program. M^{dp} is non-recursive, which we will show by writing out the program $\mathtt{prog}_{M^{\mathsf{dp}}}$ that it suggests:

$$\mathtt{def}\ \mathtt{prog}_{M^{\mathsf{dp}}}(R_1, \ldots, R_n): \tag{87}$$
$$D_1 := R_1 \tag{88}$$
$$D_2 := R_2 \tag{89}$$
$$\vdots \tag{90}$$
$$D_n := R_n \tag{91}$$
$$D := \langle D_1, D_2, \ldots, D_n \rangle \tag{92}$$
$$O := \mathcal{A}(D) \tag{93}$$

More formally, let $[\![M]\!](\vec{x}).\vec{Y}$ be the joint distribution over values for the variables $\vec{Y}$ that results from the background variables $\vec{X}$ taking on the values $\vec{x}$ (where these vectors use the same ordering). That is, $[\![M]\!](\vec{x}).\vec{Y}(\vec{y})$ represents the probability of $\vec{Y} = \vec{y}$ given that the background variables had values $\vec{X} = \vec{x}$. Since the SEM is non-recursive this can be calculated in a bottom up fashion. For example,

$$[\![M^{\mathsf{dp}}]\!](r_1,\ldots,r_n).R_i(r_i) = 1 \tag{94}$$
$$[\![M^{\mathsf{dp}}]\!](r_1,\ldots,r_n).D_i(r_i) = \Pr_{F_{D_i}}[F_{D_i}(R_i){=}r_i] = \Pr_{F_{D_i}}[R_i{=}r_i] = 1 \tag{95}$$
$$[\![M^{\mathsf{dp}}]\!](r_1,\ldots,r_n).D(\langle r_1,\ldots,r_n\rangle) = \Pr_{F_D}[F_D(D_1,\ldots,D_n){=}\langle r_1,\ldots,r_n\rangle] \tag{96}$$
$$= \Pr_{F_D}[F_D(F_{D_1}(R_1),\ldots,F_{D_n}(R_n)){=}\langle r_1,\ldots,r_n\rangle] \tag{97}$$
$$= \Pr_{F_D}[F_D(R_1,\ldots,R_n){=}\langle r_1,\ldots,r_n\rangle] \tag{98}$$
$$= \Pr_{F_D}[\langle R_1,\ldots,R_n\rangle{=}\langle r_1,\ldots,r_n\rangle] = 1 \tag{99}$$
$$[\![M^{\mathsf{dp}}]\!](r_1,\ldots,r_n).O(o) = \Pr_{F_O}[F_O(D){=}o] = \Pr_{\mathcal{A}}[\mathcal{A}(\langle r_1,\ldots,r_n\rangle){=}o] \tag{100}$$

Let a *probabilistic SEM* $\langle M, \mathcal{P}\rangle$ also have a probability distribution $\mathcal{P}$ over the background variables. We can raise the calculations above to work over $\mathcal{P}$ instead of a concrete assignment of values $\vec{x}$. Intuitively, the only needed change is that, for background variables $\vec{X}$,

$$\Pr_{\langle M,\mathcal{P}\rangle}[\vec{Y}{=}\vec{y}] = \sum_{\vec{x}\in\vec{\mathcal{X}}} \Pr_{\mathcal{P}}[\vec{X}{=}\vec{x}] * [\![M]\!](\vec{x}).\vec{Y}(\vec{y}) \tag{101}$$

where $\vec{X}$ are all the background variables.[2]

Let M be an SEM, Y be an endogenous variable of M, and y be a value that Y can take on. Pearl defines the *sub-model* $M[Z{:=}z]$ to be the SEM that results from replacing the equation $Z := F_Z(\vec{Z})$ in $\mathcal{E}$ of M with the equation $Z := z$. The sub-model $M[Z{:=}z]$ shows the *effect* of setting Z to z. Let $\Pr_{\langle M,\mathcal{P}\rangle}[Y{=}y \mid \mathrm{do}(Z{:=}z)]$ be $\Pr_{\langle M[Z:=z],\mathcal{P}\rangle}[Y{=}y]$. Note that is this well defined even when $\Pr_{\langle M,\mathcal{P}\rangle}[Z{=}z] = 0$ as long as z is within in the range of values $\mathcal{Z}$ that Z can take on.

The following lemma will not only be useful, but will illustrate the above general points on the model M^{dp} that concerns us.

[2] This is Pearl's equation (7.2) raised to work on probabilistic structural equations F_V [11, p. 205].

Lemma 2. For all $\mathcal{P}$, all o, and all $d_1,\ldots,d_n$,

$$\Pr_{\langle M^{\mathsf{dp}},\mathcal{P}\rangle}[O{=}o \mid \mathrm{do}(D_1{:=}d_1,\ldots,D_n{:=}d_n)] = \Pr_{\mathcal{A}}[\mathcal{A}(d_1,\ldots,d_n){=}o] \tag{102}$$

Proof. Let $F_{d_i}()$ represent the constant function with no arguments that always returns d_i. The structural equation for D_i is F_{d_i} in $M^{\mathsf{dp}}[D_1{:=}d_1]\cdots[D_n{:=}d_n]$. As before, we compute bottom up, but this time on the modified SEM:

$$[\![M^{\mathsf{dp}}[D_1{:=}d_1]\cdots[D_n{:=}d_n]]\!](r_1,\ldots,r_n).R_i(r_i) = 1 \tag{103}$$
$$[\![M^{\mathsf{dp}}[D_1{:=}d_1]\cdots[D_n{:=}d_n]]\!](r_1,\ldots,r_n).D_i(d_i) = \Pr_{F_{d_i}}[F_{d_i}(){=}d_i] = 1 \tag{104}$$
$$[\![M^{\mathsf{dp}}[D_1{:=}d_1]\cdots[D_n{:=}d_n]]\!](r_1,\ldots,r_n).D(\langle d_1,\ldots,d_n\rangle) = \Pr_{F_D}[F_D(D_1,\ldots,D_n){=}\langle d_1,\ldots,d_n\rangle] \tag{105}$$
$$= \Pr_{F_D}[F_D(F_{D_1}(),\ldots,F_{D_n}()){=}\langle d_1,\ldots,d_n\rangle] \tag{106}$$
$$= \Pr_{F_D}[F_D(d_1,\ldots,d_n){=}\langle d_1,\ldots,d_n\rangle] \tag{107}$$
$$= \Pr_{F_D}[\langle d_1,\ldots,d_n\rangle{=}\langle d_1,\ldots,d_n\rangle] = 1 \tag{108}$$
$$[\![M^{\mathsf{dp}}[D_1{:=}d_1]\cdots[D_n{:=}d_n]]\!](r_1,\ldots,r_n).O(o) = \Pr_{F_O}[F_O(D){=}o] = \Pr_{\mathcal{A}}[\mathcal{A}(\langle d_1,\ldots,d_n\rangle){=}o] \tag{109}$$

Thus,

$$\Pr_{\langle M^{\mathsf{dp}},\mathcal{P}\rangle}[O{=}o \mid \mathrm{do}(D_1{:=}d_1,\ldots,D_n{:=}d_n)] = \Pr_{\langle M^{\mathsf{dp}}[D_1:=d_1]\cdots[D_n:=d_n],\mathcal{P}\rangle}[O{=}o] \tag{110}$$

$$= \sum_{\vec{r}\in\mathcal{R}^n} \Pr_{\mathcal{P}}[\vec{R}{=}\vec{r}] * [\![M^{\mathsf{dp}}[D_1{:=}d_1]\cdots[D_n{:=}d_n]]\!](\vec{r}).O(o) \tag{111}$$

$$= \sum_{\vec{r}\in\mathcal{R}^n} \Pr_{\mathcal{P}}[\vec{R}{=}\vec{r}] * \Pr_{\mathcal{A}}[\mathcal{A}(\langle d_1,\ldots,d_n\rangle){=}o] \tag{112}$$

$$= \Pr_{\mathcal{A}}[\mathcal{A}(\langle d_1,\ldots,d_n\rangle){=}o] * \textstyle\sum_{\vec{r}\in\mathcal{R}^n} \Pr_{\mathcal{P}}[\vec{R}{=}\vec{r}] \tag{113}$$

$$= \Pr_{\mathcal{A}}[\mathcal{A}(\langle d_1,\ldots,d_n\rangle){=}o] * 1 \tag{114}$$

$$= \Pr_{\mathcal{A}}[\mathcal{A}(\langle d_1,\ldots,d_n\rangle){=}o] \tag{115}$$

□

Lemma 3. For all $\mathcal{P}$, all o, all $r_1,\ldots,r_n$, and all $d_1,\ldots,d_n$,

$$\Pr_{\langle M^{\mathsf{dp}},\mathcal{P}\rangle}[O{=}o \mid \mathrm{do}(D_j{=}d_j)] \tag{116}$$

$$= \sum_{\langle r_1,\ldots,r_{j-1},r_{j+1},\ldots,r_n\rangle\in\mathcal{R}^{n-1}} \Pr_{\mathcal{P}}\left[\wedge_{i\in\{1,\ldots,j-1,j+1,\ldots,n\}} R_i{=}r_i\right] * \Pr_{\mathcal{A}}[\mathcal{A}(r_1,\ldots,r_{j-1},d_j,r_{j+1},\ldots,r_n){=}o] \tag{117}$$

$$= \sum_{\langle d_1,\ldots,d_{j-1},d_{j+1},\ldots,d_n\rangle\in\mathcal{D}^{n-1}} \Pr_{\langle M^{\mathsf{dp}},\mathcal{P}\rangle}\left[\wedge_{i\in\{1,\ldots,j-1,j+1,\ldots,n\}} D_i{=}d_i\right] * \Pr_{\mathcal{A}}[\mathcal{A}(d_1,\ldots,d_n){=}o] \tag{118}$$

Proof. With out loss of generality, assume j is 1. Let $F_{d_1}()$ represent the constant function with no arguments that always returns $d_1 = d_j$. The structural equation for D_1 is F_{d_1} in $M^{\mathsf{dp}}[D_1{:=}d_1]$. As before, we compute bottom up, but this time on the modified SEM:

$$[\![M^{\mathsf{dp}}[D_1{:=}d_1]]\!](r_1,\ldots,r_n).R_i(r_i) = 1 \tag{119}$$

$$\tag{120}$$

holds before. The behavior of D_i varies based on whether $i = 1$:

$$[\![M^{\mathsf{dp}}[D_1{:=}d_1]]\!](r_1,\ldots,r_n).D_i(r_i) = \Pr_{F_{D_i}}[F_{D_i}(R_i){=}r_i] = \Pr_{F_{D_i}}[R_i{=}r_i] = 1 \qquad \text{for all } i \neq 1 \tag{121}$$

$$[\![M^{\mathsf{dp}}[D_1{:=}d_1]]\!](r_1,\ldots,r_n).D_1(d_1) = \Pr_{F_{d_1}}[F_{d_1}(){=}d_1] = 1 \tag{122}$$

Thus,

$$[\![M^{\mathsf{dp}}[D_1{:=}d_1]]\!](r_1,\ldots,r_n).D(\langle d_1,r_2,\ldots,r_n\rangle) = \Pr_{F_D}[F_D(D_1,D_2\ldots,D_n){=}\langle d_1,r_2,\ldots,r_n\rangle] \tag{123}$$

$$= \Pr_{F_D}[F_D(F_{d_1}(),F_{D_2}(R_2),\ldots,F_{D_n}(R_n)){=}\langle d_1,r_2,\ldots,r_n\rangle] \tag{124}$$

$$= \Pr_{F_D}[F_D(d_1,r_2,\ldots,r_n){=}\langle d_1,r_2,\ldots,r_n\rangle] \tag{125}$$

$$= \Pr_{F_D}[\langle d_1,r_2,\ldots,d_n\rangle{=}\langle d_1,r_2,\ldots,r_n\rangle] = 1 \tag{126}$$

$$[\![M^{\mathsf{dp}}[D_1{:=}d_1]]\!](r_1,\ldots,r_n).O(o) = \Pr_{F_O}[F_O(D){=}o] = \Pr_{\mathcal{A}}[\mathcal{A}(\langle d_1,r_2,\ldots,r_n\rangle){=}o] \tag{127}$$

Thus,

$$\Pr_{\langle M^{\mathsf{dp}},\mathcal{P}\rangle}[O{=}o \mid \mathrm{do}(D_1{:=}d_1)] \tag{128}$$

$$= \Pr_{\langle M^{\mathsf{dp}}[D_1:=d_1],\mathcal{P}\rangle}[O{=}o] \tag{129}$$

$$= \sum_{r_1,\ldots,r_n\in\mathcal{R}^n} \Pr_{\mathcal{P}}[R_1{=}r_1,\ldots,R_n{=}r_n] * [\![M^{\mathsf{dp}}[D_1:=d_1]]\!](r_1,\ldots,r_n).O(o) \tag{130}$$

$$= \sum_{r_1,\ldots,r_n\in\mathcal{R}^n} \Pr_{\mathcal{P}}[R_1{=}r_1,\ldots,R_n{=}r_n] * \Pr_{\mathcal{A}}[\mathcal{A}(\langle d_1,r_2,\ldots,r_n\rangle){=}o] \tag{131}$$

$$= \sum_{r_1,\ldots,r_n\in\mathcal{R}^n} \Pr_{\mathcal{P}}[R_1{=}r_1 \mid R_2{=}r_2,\ldots,R_n{=}r_n] * \Pr_{\mathcal{P}}[R_2{=}r_2,\ldots,R_n{=}r_n] * \Pr_{\mathcal{A}}[\mathcal{A}(\langle d_1,r_2,\ldots,r_n\rangle){=}o] \tag{132}$$

$$= \sum_{r_2,\ldots,r_n\in\mathcal{R}^n} \sum_{r_1\in\mathcal{R}} \Pr_{\mathcal{P}}[R_1{=}r_1 \mid R_2{=}r_2,\ldots,R_n{=}r_n] * \Pr_{\mathcal{P}}[R_2{=}r_2,\ldots,R_n{=}r_n] * \Pr_{\mathcal{A}}[\mathcal{A}(\langle d_1,r_2,\ldots,r_n\rangle){=}o] \tag{133}$$

$$= \sum_{r_2,\ldots,r_n\in\mathcal{R}^n} \Pr_{\mathcal{P}}[R_2{=}r_2,\ldots,R_n{=}r_n] * \Pr_{\mathcal{A}}[\mathcal{A}(\langle d_1,r_2,\ldots,r_n\rangle){=}o] * \textstyle\sum_{r_1\in\mathcal{R}} \Pr_{\mathcal{P}}[R_1{=}r_1 \mid R_2{=}r_2,\ldots,R_n{=}r_n] \tag{134}$$

$$= \sum_{r_2,\ldots,r_n\in\mathcal{R}^n} \Pr_{\mathcal{P}}[R_2{=}r_2,\ldots,R_n{=}r_n] * \Pr_{\mathcal{A}}[\mathcal{A}(\langle d_1,r_2,\ldots,r_n\rangle){=}o] * 1 \tag{135}$$

$$= \sum_{r_2,\ldots,r_n\in\mathcal{R}^n} \Pr_{\mathcal{P}}[R_2{=}r_2,\ldots,R_n{=}r_n] * \Pr_{\mathcal{A}}[\mathcal{A}(\langle d_1,r_2,\ldots,r_n\rangle){=}o] \tag{136}$$

$$= \sum_{d_2,\ldots,d_n\in\mathcal{D}^n} \Pr_{\mathcal{P}}[D_2{=}d_2,\ldots,D_n{=}d_n] * \Pr_{\mathcal{A}}[\mathcal{A}(\langle d_1,d_2,\ldots,d_n\rangle){=}o] \tag{137}$$

where the last line follows since $D_i = R_i$ for $i \neq 1$. □

References

[1] D. Kifer and A. Machanavajjhala, "No free lunch in data privacy," in *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data*. ACM, 2011, pp. 193–204.

[2] ——, "A rigorous and customizable framework for privacy," in *Proceedings of the 31st ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*, ser. PODS '12. ACM, 2012, pp. 77–88.

[3] X. He, A. Machanavajjhala, and B. Ding, "Blowfish privacy: Tuning privacy-utility trade-offs using policies," in *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD 2014)*. ACM, 2014.

[4] D. Kifer and A. Machanavajjhala, "Pufferfish: A framework for mathematical privacy definitions," *ACM Trans. Database Syst.*, vol. 39, no. 1, pp. 3:1–3:36, 2014.

[5] C. Liu, S. Chakraborty, and P. Mittal, "Dependence makes you vulnerable: Differential privacy under dependent tuples," in *Network and Distributed System Security Symposium (NDSS)*. The Internet Society, 2016.

[6] R. Bassily, A. Groce, J. Katz, and A. Smith, "Coupled-worlds privacy: Exploiting adversarial uncertainty in statistical data privacy," in *Proceedings of the 2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, ser. FOCS '13. IEEE Computer Society, 2013, pp. 439–448.

[7] S. P. Kasivisiwanathan and A. Smith, "On the 'semantics' of differential privacy: A bayesian formulation," *Journal of Privacy and Confidentiality*, vol. 6, no. 1, pp. 1–16, 2014.

[8] F. McSherry, "Lunchtime for data privacy," Blog: https://github.com/frankmcsherry/blog/blob/master/posts/2016-08-16.md, 2016.

[9] ——, "Differential privacy and correlated data," Blog: https://github.com/frankmcsherry/blog/blob/master/posts/2016-08-29.md, 2016.

[10] C. Dwork, F. Mcsherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of Cryptography Conference*. Springer, 2006, pp. 265–284.

[11] J. Pearl, *Causality*, 2nd ed. Cambridge University Press, 2009.

[12] C. Dwork, "Differential privacy," in *Automata, Languages and Programming, 33rd International Colloquium, ICALP 2006, Venice, Italy, July 10–14, 2006, Proceedings, Part II*, ser. Lecture Notes in Computer Science, M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener, Eds., vol. 4052. Springer, 2006, pp. 1–12.

[13] G. Smith, "Recent developments in quantitative information flow (invited tutorial)," in *Proceedings of the 2015 30th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, ser. LICS '15. IEEE Computer Society, 2015, pp. 23–31.

[14] J. Tang, A. Korolova, X. Bai, X. Wang, and X. Wang, "Privacy loss in Apple's implementation of differential privacy on MacOS 10.12," *CoRR*, vol. 1709.02753, 2017.

[15] R. A. Fisher, *The Design of Experiments*. Oliver & Boyd, 1935.

[16] M. C. Tschantz, A. Datta, A. Datta, and J. M. Wing, "A methodology for information flow experiments," in *Computer Security Foundations Symposium*. IEEE, 2015.

[17] A. Datta, M. C. Tschantz, and A. Datta, "Automated experiments on ad privacy settings: A tale of opacity, choice, and discrimination," in *Proceedings on Privacy Enhancing Technologies (PoPETs)*. De Gruyter Open, 2015.

[18] F. McSherry, "Privacy integrated queries," in *Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data (SIGMOD)*. Association for Computing Machinery, Inc., 2009.

[19] S. L. Warner, "Randomized response: A survey technique for eliminating evasive answer bias," *Journal of the American Statistical Association*, vol. 60, no. 309, pp. 63–69, 1965. [Online]. Available: http://www.jstor.org/stable/2283137

[20] T. Dalenius, "Towards a methodology for statistical disclosure control," *Statistik Tidskrift*, vol. 15, pp. 429–444, 1977.

[21] C. Dwork and M. Naor, "On the difficulties of disclosure prevention in statistical databases or the case for differential privacy," *Journal of Privacy and Confidentiality*, vol. 2, no. 1, pp. 93–107, 2008.

[22] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, "Our data, ourselves: Privacy via distributed noise generation," in *Proceedings of the 24th Annual International Conference on The Theory and Applications of Cryptographic Techniques*, ser. EUROCRYPT'06. Springer-Verlag, 2006, pp. 486–503.

[23] S. P. Kasiviswanathan and A. D. Smith, "A note on differential privacy: Defining resistance to arbitrary side information," *CoRR*, vol. 0803.3946, 2008.

[24] G. Cormode, "Personal privacy vs population privacy: Learning to attack anonymization," in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '11. ACM, 2011, pp. 1253–1261.

[25] J. Gehrke, E. Lui, and R. Pass, "Towards privacy for social networks: A zero-knowledge based definition of privacy," in *Proceedings of the 8th Conference on Theory of Cryptography*, ser. TCC'11. Springer-Verlag, 2011, pp. 432–449.

Colophon

The authors typeset this document using LaTeX with the `tufte-handout` document class. We configured the class with the `nobib`, `nofonts`, and `justified` options. We also altered the appearance of section headers.

The varying line widths in the document are a purposeful attempt at balancing two competing concerns in typesetting. On the one hand, people find reading long lines of text difficult, which argues for short line lengths. On the other hand, series of equations are easier to follow when the individual equations are not broken up across lines, which argues for line lengths long enough to hold the longest equation. To balance these two concerns, we use a short line length for text, but exceed that length as needed for wide equations.

This compromise sacrifices the consistency of line lengths. We welcome comments on whether this sacrifice is too high a price to pay for balancing the two aforementioned concerns.