Towards an Axiomatization of Statistical Privacy and Utility

Penn State University Technical Report #CSE-10-002

Daniel Kifer
Penn State University

Bing-Rong Lin
Penn State University

ABSTRACT

"Privacy" and "utility" are words that frequently appear in the literature on statistical privacy. But what do these words really mean? In recent years, many problems with intuitive notions of privacy and utility have been uncovered. Thus more formal notions of privacy and utility, which are amenable to mathematical analysis, are needed. In this paper we present our initial work on an axiomatization of privacy and utility. In particular, we study how these concepts are affected by randomized algorithms. Our analysis yields new insights into the construction of both privacy definitions and mechanisms that generate data according to such definitions. In particular, it characterizes a class of relaxations of differential privacy and shows that desirable outputs of a differentially private mechanism are best interpreted as certain graphs rather than query answers or synthetic data.

1. INTRODUCTION

Statistical privacy is the art of designing a privacy mechanism that transforms sensitive data into data that are simultaneously useful and non-sensitive. The sensitive data typically contain private information about individuals (e.g., income, medical history, search queries) or organizations (e.g., intranet network traces, customer records) and are usually collected by businesses (e.g., Netflix, AOL) or government agencies (e.g., U.S. Census Bureau).

Non-sensitive data produced by privacy mechanisms are highly desirable because they can be made available to the public without restrictions on access. Researchers will benefit from previously unavailable data – they could, for example, study socio-economic and business trends, develop new models, and design and evaluate new algorithms using such data.

All of this potential success hinges on two poorly-defined words: *privacy* and *utility*. What does it mean for a privacy mechanism to output a dataset that is non-sensitive? What does it mean for a privacy mechanism to output a dataset that has high utility (i.e. is useful)? The literature is full of definitions of what privacy is and is not; it is also full of ways of assigning a numerical score to the utility of a dataset (for recent surveys, see [10, 21]).

However, current privacy definitions and utility measures are typically constructed on the basis of intuition, but intuition alone can lead us astray. Some spectacular privacy breaches (such as demonstrations involving AOL [3] and GIC [39] data) have occurred when such intuition was not followed by a thorough analysis. In other cases, subtle implicit assumptions created weaknesses that could be exploited to breach privacy [25, 40, 22, 26]. Similarly, the choice of a privacy mechanism based on some intuitively plausible measures of utility can result in a dataset that is not as useful as it could be [31, 23]. For example, Ghosh et al. [23] have shown that if utility is measured by expected loss (in the Bayesian sense) then it is possible that a "suboptimal" privacy mechanism followed by a lossy postprocessing step can mimic an "optimal" privacy mechanism, thus casting doubts on the appropriateness of expected loss.

Clearly, a unified theory of privacy and utility is needed to guide the development of privacy definitions, utility measures, and privacy mechanisms. We believe that the path to such a theory relies on an axiomatization of privacy and utility. That is, we must examine axioms for what privacy and utility should mean and then study the consequences of those axioms. When new sensitive data need to be released, a data publisher can pick and choose whatever axioms are appropriate for the application at hand. The data publisher can then select an appropriate privacy mechanism for generating non-sensitive data that are safe for release.

The benefit of this approach is that a small set of axioms can be thoroughly studied but a large, disjointed set of privacy definitions and utility measures cannot. Intuitions would be formally justified (or discredited) by axioms, which would then serve as explanations for why some intuition should or should not be followed. Therefore with an axiomatic approach we can reduce the possibility of privacy breaches and useless datasets caused by faulty intuition.

In this paper we present some of our work on such a theory. In fact, our axioms lead to several concrete results. The first main result answers questions about how differential privacy [14] can be relaxed. Differential privacy is a formal (and very stringent) privacy definition that uses a set of predicates to restrict the output probabilities of a privacy mechanism. Relaxations of differential privacy are studied as a way of improving the utility of data that are

A conference proceedings version will appear in PODS 2010.

output from privacy mechanisms (e.g., [15, 32, 29]). These relaxations frequently change the predicates that differential privacy uses. In this paper we characterize the *class* of predicates that can be used (instead of just presenting one or two relaxed definitions). The result is Definition 2.1.3. Thus we shift the question from "how can differential privacy be relaxed" to "how can we design privacy mechanisms to take advantage of these relaxations".

Our second main contribution deals with utility. For differential privacy, we answer the question of what does a desirable privacy mechanism look like. We show that the outputs of such a mechanism do not necessarily correspond to query answers or even data that have the same format as the original sensitive data. Instead, the outputs can be viewed as a certain collection of trees which can be interpreted as likelihood functions. We then relate this to the famous likelihood principle of statistics [9]. We also discuss what desirable utility measures should look like and prove that a privacy mechanism called the Geometric Mechanism [23] satisfies one such utility measure which is coincidentally used in the study of Markov chains.

Our results hinge on three axioms that deal with the effects of randomized algorithms on privacy and utility. Informally, the first axiom states that post-processing the output of a privacy mechanism should not decrease privacy (for example, subsampling nonsensitive data should yield nonsensitive data); the second axiom states that a random choice from a set of privacy mechanisms is at least as good as the weakest of those mechanisms. These two axioms enforce a certain internal consistency for privacy definitions. The third axiom states that utility is an intrinsic property of a dataset so that postprocessing cannot increase the amount of information it contains (for example, applying a non-invertible transformation should strictly decrease utility). We also comment on when these axioms may or may not be appropriate.

The rest of this paper is organized as follows. Our main results on privacy are discussed in Section 2. We present our privacy axioms in Section 2.1, where we also give an overview of our technical results about the consequences of those axioms. We give necessary conditions for a generalization of differential privacy to satisfy these axioms in Section 2.2 and we give sufficient conditions in Section 2.3. Utility is discussed in Section 3. We present a utility axiom in Section 3.1. We give examples of appropriate and inappropriate measures of utility in Section 3.2. We then characterize some optimal differentially private mechanisms in Section 3.3.

2. REASONING ABOUT PRIVACY

In this section, we present our results on some privacy axioms and their consequences. First we present some basic definitions, including *abstract differential privacy* (Definition 2.0.3) which is an abstract version of differential privacy. In Section 2.1 we present two privacy axioms and give an overview of our main technical results that lead to a concrete generalization (and relaxation) of differential privacy known as *generic differential privacy* (Definition 2.1.3). In Section 2.2 we present our technical results that use our privacy axioms to characterize abstract differential privacy in terms of necessary conditions. In Section 2.3 we present a proof of sufficient conditions. These results lead from abstract differential privacy to generic differential privacy.

In order to study how randomized algorithms affect privacy and utility, we need to formalize what we mean by a "randomized algorithm". Here we are more interested in the probabilistic aspects than in the computational aspects. To avoid unnecessary distinctions between distributions that are discrete, finite, or a mixture of the two, we will use the language of measure theory [36] in the way it is commonly used in the statistics literature [38].

Definition 2.0.1. (Randomized Algorithm). *Given an input space $\mathbb{I}$ with associated σ-algebra $\mathbb{S}_I$ and probability measure μ, and an output space $\mathbb{O}$ with a σ-algebra $\mathbb{S}_O$, a randomized algorithm $\mathcal{A}$ is a measurable function from $\mathbb{I}$ to $\mathbb{O}$ such that the induced conditional probability $P_{\mathcal{A}}(O \mid I)$ (for $O \in \mathbb{S}_O$ and $I \in \mathbb{S}_I$) is a regular conditional probability*[1].

For readers who are unacquainted with measure theory, it is generally safe (i.e. barring extremely pathological cases) to interpret Definition 2.0.1 as saying that a randomized algorithm is a conditional probability distribution that specifies the probability of an output $o \in \mathbb{O}$ given an input $i \in \mathbb{I}$. Note that a randomized algorithm may be completely deterministic.

It is important to note that each input $i \in \mathbb{I}$ corresponds to a possible **dataset** and *not* to a **tuple** in a dataset. An output $o \in \mathbb{O}$ could be anything – a set of query answers, synthetic data, or some other object. Thus all of the randomized algorithms we consider here take a dataset as an input and they output some object $o \in \mathbb{O}$. In particular, they capture all possible processes that create sanitized data.

Composition of two randomized algorithms $\mathcal{A}_1$ and $\mathcal{A}_2$ is denoted by $\mathcal{A}_1 \circ \mathcal{A}_2$ and is defined as the application of $\mathcal{A}_2$ followed by $\mathcal{A}_1$ (assuming the output space of $\mathcal{A}_2$ is the same as the input space of $\mathcal{A}_1$). The resulting conditional distribution $P(Z|x)$ is then $\int P_{\mathcal{A}_1}(Z|y)P_{\mathcal{A}_2}(dy|x)$ (or $\int P_{\mathcal{A}_1}(Z|y)P_{\mathcal{A}_2}(y|x)\ dy$ for those unfamiliar with measure theory; for discrete random variables replace the integral with a sum).

In this paper we are considering the scenario where a *data publisher* possesses sensitive information about individuals. The data publisher would like to release some version of this data without violating the privacy of those individuals. An *attacker* (or a class of attackers) will try to infer the sensitive information from the released data. The data publisher first selects a privacy definition that would defend against a certain class of attackers. Then the data publisher selects a special randomized algorithm known as a *privacy mechanism*, denoted by $\mathfrak{M}$, which satisfies the privacy definition. Finally, the data publisher applies the privacy mechanism $\mathfrak{M}$ to the sensitive data, and releases the output of $\mathfrak{M}$. We will refer to the output of $\mathfrak{M}$ as *sanitized data* to emphasize the fact that it should be safe to release to the public. Note that we will use the symbol $\mathfrak{M}$ to refer to any randomized algorithm that is a privacy mechanism and $\mathcal{A}$ to refer to a randomized algorithm in general.

The privacy axioms that we will discuss in Section 2.1 are not tied to any specific privacy definition. However, we will use those axioms to add insight to the definition known as differential privacy; in particular, they will show how we can generalize and relax its stringent conditions. Thus we discuss differential privacy next.

Definition 2.0.2. (Differential Privacy [14]). *Let $\mathbb{I}$ be a set of database instances and $\epsilon > 0$. A randomized algorithm*

[1]i.e. $P(O|i)$ is a probability measure for each fixed $i \in \mathbb{I}$ and is a measurable function of i for each fixed $O \in \mathbb{S}_O$

$\mathfrak{M}$ with output space $\mathbb{O}$ satisfies ϵ-differential privacy if for all measurable $O \subseteq \mathbb{O}$ and for all pairs (i_1, i_2) of database instances that differ only in the insertion or deletion of one individual's information, $P_{\mathfrak{M}}(O \mid i_1) \leq e^{\epsilon} P_{\mathfrak{M}}(O \mid i_2)$.

We will refer to $\mathbb{I}$ as the input space and $\mathbb{O}$ as the output space. Our first step is to introduce the notion of a *privacy relation* $\mathcal{R}$, which is an irreflexive binary relation $\subseteq \mathbb{I} \times \mathbb{I}$. $\mathcal{R}$ generalizes the notion of neighboring databases in that we will require $P_{\mathfrak{M}}(O \mid i_1) \leq e^{\epsilon} P_{\mathfrak{M}}(O \mid i_2)$ only for $(i_1, i_2) \in \mathcal{R}$. We will also replace the condition $P_{\mathfrak{M}}(O \mid i_1) \leq e^{\epsilon} P_{\mathfrak{M}}(O \mid i_2)$ with conditions of the form

$$\mathrm{q}_{i_1,i_2}(P_{\mathfrak{M}}(O \mid i_1), P_{\mathfrak{M}}(O \mid i_2)) = T$$

where $\{\mathrm{q}_{i_1,i_2}\}_{(i_1,i_2)\in\mathcal{R}}$ is an arbitrary set of predicates, which we call the *privacy predicates*. Having multiple predicates allows us to customize the privacy definition based on the sensitivity of each possible dataset $i \in \mathbb{I}$; a region of datasets in $\mathbb{I}$ containing little private information could use less stringent predicates. In contrast, the traditional versions of differential privacy assume that all possible database instances have the same privacy requirements. This leads to the following privacy definition:

Definition 2.0.3. (Abstract Differential Privacy). *Suppose we have an input space $\mathbb{I}$, output space $\mathbb{O}$, a binary irreflexive relation $\mathcal{R} \subseteq \mathbb{I} \times \mathbb{I}$, and a binary predicate $\mathrm{q}_{i_1,i_2} : [0,1] \times [0,1] \to \{T, F\}$ for each $(i_1, i_2) \in \mathcal{R}$. A randomized algorithm $\mathfrak{M}$ satisfies abstract differential privacy for $\{\mathrm{q}_{i_1,i_2}\}_{(i_1,i_2)\in\mathcal{R}}$ if for all measurable $O \subseteq \mathbb{O}$ and for all $(i_1, i_2) \in \mathcal{R}$ we must have $\mathrm{q}_{i_1,i_2}(P_{\mathfrak{M}}(O|i_1), P_{\mathfrak{M}}(O|i_2)) = T$.*

This abstraction serves two purposes. First, it allows us to take an approach similar to [30] where we can avoid assigning inessential semantic information to the input or output spaces. For example, the arity of the schemas for the instances in $\mathbb{I}$ and the number of tuples containing an individual's information are irrelevant except for their effect on the topology of the privacy relation $\mathcal{R}$ when viewed as a directed graph.

The second purpose of this abstraction is to study the essential properties of the privacy predicates q_{i_1,i_2}. One generalization already exists: the condition that $P_{\mathfrak{M}}(O \mid i_1) \leq e^{\epsilon} P_{\mathfrak{M}}(O \mid i_2) + \delta$, for some small δ [15, 32]. What other predicates can be used, what do they look like, and what are their properties? These questions are answered in Sections 2.2 and 2.3 which characterize the class of such predicates. The privacy axioms which form the foundation for these results are discussed next in Section 2.1, which also contains an informal overview of our main results.

2.1 Privacy Axioms

What makes a good privacy definition and how should the data publisher choose one? We feel that this question must be addressed axiomatically. In general, a data publisher would select the axioms that are appropriate to the application at hand. The two axioms we present here are designed to enforce a certain internal consistency for privacy definitions. Our first axiom, Axiom 2.1.1 deals with the effects of postprocessing the sanitized data (this axiom has been observed to hold in differential privacy [14, 24, 2, 41], but we do not tie it to any specific privacy definition).

Axiom 2.1.1. (Transformation Invariance). *Suppose we have a privacy definition, a privacy mechanism $\mathfrak{M}$ that satisfies this definition, and a randomized algorithm $\mathcal{A}$ whose input space is the output space of $\mathfrak{M}$ and whose randomness is independent of both the data and the randomness in $\mathfrak{M}$. Then $\mathfrak{M}' \equiv \mathcal{A} \circ \mathfrak{M}$ must also be a privacy mechanism satisfying that privacy definition.*

Essentially this axiom says that postprocessing sanitized data maintains privacy as long as the postprocessing algorithm does not use the sensitive information directly (i.e. sensitive information is only used indirectly via the sanitized data).

Note that this axiom is very strong in some ways – it places no computational restrictions on $\mathcal{A}$ and so encrypting a database (for example, with DES) would not qualify as a privacy mechanism. If a data publisher feels that this axiom is too strong for the application at hand, it can be replaced with some form of invariance with respect to a subset of randomized algorithms. On the other hand, a strengthening of the axiom can discuss an attacker's prior knowledge about the data.This may allow a formalization of k-anonymity [37, 39], ℓ-diversity [28], and related definitions, but such variations of the axiom are outside of the scope of this paper.

We shall also make use of the following axiom.

Axiom 2.1.2. (Privacy Axiom of Choice) *Given a privacy definition, let $\mathfrak{M}_1$ and $\mathfrak{M}_2$ be privacy mechanisms that satisfy the privacy definition. For any $p \in [0,1]$, let $\mathfrak{M}_p$ be a randomized algorithm that on input i outputs $\mathfrak{M}_1(i)$ with probability p (independent of the data and the randomness in $\mathfrak{M}_1$ and $\mathfrak{M}_2$) and $\mathfrak{M}_2(i)$ with probability $1-p$. Then $\mathfrak{M}_p$ is a privacy mechanism that satisfies the privacy definition.*

This axiom allows us to randomly pick a privacy mechanism, as long as our decision is not influenced by the data. We believe that this is a fundamental axiom that should be required for any application of statistical privacy.

We next present an overview of our main results and show how they lead to a generic version of differential privacy that satisfies our axioms (Definition 2.1.3). Consider again the definition of abstract differential privacy (Definition 2.0.3). Fix a pair of datasets that are neighbors according to the privacy relation, $(i_1, i_2) \in \mathcal{R}$ and consider the corresponding privacy predicate q_{i_1,i_2}. Recall that a privacy mechanism $\mathfrak{M}$ for this definition must satisfy the condition $\mathrm{q}_{i_1,i_2}(P_{\mathfrak{M}}(O \mid i_1), P_{\mathfrak{M}}(O \mid i_2)) = T$ for all measurable O. Our main technical results (Theorem 2.2.5 and 2.3.1 from Sections 2.2 and 2.3) that follow from these axioms state that q_{i_1,i_2} cannot be arbitrary, and in fact can be characterized by an upper bound function $M_{i_1,i_2}[0,1] \to [0,1]$ and a lower bound function $m_{i_1,i_2}[0,1] \to [0,1]$ in the following way. $M_{i_1,i_2}(a) > b > m_{i_1,i_2}(a)$ implies $\mathrm{q}_{i_1,i_2}(a,b) = T$ while $b > M_{i_1,i_2}(a)$ or $b < m_{i_1,i_2}(a)$ implies $\mathrm{q}_{i_1,i_2}(a,b) = F$. Furthermore, M_{i_1,i_2} is concave, $M_{i_1,i_2}(1) = 1$, it is continuous everywhere except possibly at 0, and $m_{i_1,i_1}(a) = 1 - M_{i_1,i_2}(1-a)$. Note that we did not specify what happens at the boundary $b = m_{i_1,i_2}(a)$ or $b = M_{i_1,i_2}(a)$ – our results can be strengthened to characterize this boundary as well, but only at the cost of making our presentation much more complicated (thus we have decided only to present this simplified case).

On the other hand, any such M_{i_1,i_2} and m_{i_1,i_2} can be used to define a predicate q_{i_1,i_2} such that $\mathrm{q}_{i_1,i_2}(a,b) = T \Leftrightarrow$

$M_{i_1,i_2}(a) \geq b \geq m_{i_1,i_2}(a)$ and the use of these predicates in abstract differential privacy satisfies Axioms 2.2.5 and 2.3.1.

Thus we have both necessary and sufficient conditions. Note that these are intuitively pleasing results and now they can be justified as the consequences of the two axioms presented here and without any further use of intuition.

By the properties of M_{i_1,i_2} and m_{i_1,i_2}, $M_{i_1,i_2}(a) \geq b \geq m_{i_1,i_2}(a)$ is true if and only if $M_{i_1,i_2}(a) \geq b$ and $M_{i_1,i_2}(1-a) \geq 1-b$ and so we can use this observation to define generic differential privacy:

DEFINITION 2.1.3. (Generic Differential Privacy) *Let $\mathbb{I}$ be an input space, $\mathbb{O}$ an output space, and $\mathcal{R} \subseteq \mathbb{I} \times \mathbb{I}$ a binary irreflexive relation. For each $(i_1, i_2) \in \mathcal{R}$ let $M_{i_1,i_2} : [0,1] \to [0,1]$ be concave function, continuous on $(0,1]$, with $M_{i_1,i_2}(1) = 1$. A randomized algorithm $\mathfrak{M}$ satisfies generic differential privacy if for all measurable $O \subseteq \mathbb{O}$ and for all $(i_1, i_2) \in \mathcal{R}$ we have $M_{i_1,i_2}(P_{\mathfrak{M}}(O|i_1)) \geq P_{\mathfrak{M}}(O|i_2))$ and $M_{i_1,i_2}(P_{\mathfrak{M}}(O^c|i_1)) \geq P_{\mathfrak{M}}(O^c|i_2))$, where O^c is the complement of O.*

For differential privacy, the function $M_{i_1,i_2}(a,b)$ is the same for all i_1 and i_2 and is equal to $\min(e^{\epsilon}a,\ 1-e^{-\epsilon}(1-a))$. For (ϵ,δ)-indistinguishability [15, 32], the function $M_{i_1,i_2}(a,b)$ is the same for all i_1 and i_2 and is equal to $\min\{1,\ e^{\epsilon}a + \delta,\ 1-e^{-\epsilon}(1-a-d)\}$, where δ is very small.

It is very important to note that Definition 2.1.3 covers a wide range of privacy definitions from the very stringent (e.g. differential privacy) to very lax definitions (which we will discuss next). This allows the strength of the definition to be tailored to the application at hand (we do not believe in a one-size-fits-all philosophy). Thus this is a true generalization/relaxation as we are only requiring the internal consistency enforced by Axioms 2.1.1 and 2.1.2.

One particularly lax definition results when $M_{i_1,i_2} \equiv 1$. This choice of M_{i_1,i_2} allows the "identity" mechanism, which simply outputs its inputs. Is this reasonable? For some applications, it is – for example, if there exists a nondisclosure agreement or if the data simply are not sensitive. Thus a proper class of relaxations should include this special case. Another interesting example is the "subsampling" mechanism which outputs a random subset of the input data (this can happen with various choices of M_{i_1,i_2}). Again, in some applications this is acceptable - for example, subsampling is commonly used by statistical agencies [5, 10].

It turns out that for Generic Differential Privacy (Definition 2.1.3) there is a semantic interpretation to the privacy guarantees that is similar to the semantic interpretations given by Dwork et al. [16] and Ganta et al. [22] and to γ-amplification [20]. Suppose $\mathfrak{M}$ is a privacy mechanism with output space $\mathbb{O}$. Consider two database instances i_1 and i_2 such that $(i_1, i_2) \in \mathcal{R}$. For example, i_1 and i_2 may differ only on the tuples corresponding to one individual. Suppose i_1 is the true data. An attacker may have a prior belief in the probability of i_1 and i_2. We express this as the log-odds $\log(\frac{P_{attacker}(i_2)}{P_{attacker}(i_1)})$. If $\mathfrak{M}(i_1)$ outputs some $o \in \mathbb{O}$ then the attacker's log odds will become $\log(\frac{P_{attacker}(i_2 \mid o)}{P_{attacker}(i_1 \mid o)})$. Denote the difference between them as $\Delta = \log(\frac{P_{attacker}(i_2 \mid o)}{P_{attacker}(i_1 \mid o)}) - \log(\frac{P_{attacker}(i_2)}{P_{attacker}(i_1)})$. The probability that Δ takes a value x is then the probability that any bad $o \in \mathbb{O}$ is produced which changes the log-odds by x. This random variable Δ has the following behavior:

PROPOSITION 2.1.4. *Let i_1 be the true data and let $\mathfrak{M}$ be a privacy mechanism for generic differential privacy.*[2] *If $P_{attacker}(i_1) > 0$ and $P_{attacker}(i_2) > 0$ then for $\epsilon > 0$ we have $P(\Delta \geq \epsilon \mid i_1) \leq a'$ where $a' = \sup\{a > 0 : \log \frac{M_{i_1,i_2}(a)}{a} \geq \epsilon\}$. Similarly, $P(\Delta \leq -\epsilon \mid i_1) \leq a''$ where $a'' = \sup\{a > 0 : \log \frac{m_{i_1,i_2}(a)}{a} \leq -\epsilon\}$ (with the convention that $\sup \emptyset = 0$). In both cases the probability depends only on the randomness in $\mathfrak{M}$.*

The proof is in Appendix D. Intuitively, we can interpret this proposition as follows. Suppose an attacker knows everything in the database except for Bob's information and suppose Bob has cancer. Proposition 2.1.4 probabilistically bounds the increase and decrease in attacker's belief between the alternatives "Bob has cancer" vs. "Bob has flu" (or any other disease). Note that the definitions of the quantities a' and a'' in Proposition 2.1.4 make sense because our proof shows that $\log \frac{M_{i_1,i_2}(a)}{a} \geq 0$ and is nonincreasing in a while $\log \frac{m_{i_1,i_2}(a)}{a} \leq 0$ and is a nondecreasing function. Again, this is simply a consequence of our axioms.

2.2 Characterizing Abstract Differential Privacy (necessary conditions)

In this section we characterize the class of privacy predicates q_{i_1,i_2} that make Definition 2.0.3 (abstract differential privacy) satisfy Axioms 2.1.1 and 2.1.2.

Fix i_1 and $i_2 \in \mathbb{I}$ such that $(i_1, i_2) \in \mathcal{R}$. This allows us to drop the notational dependency of the privacy predicate q_{i_1,i_2} on i_1 and i_2 so that we can simply refer to it as q. Recall that if $\mathfrak{M}$ is a privacy mechanism for abstract differential privacy (Definition 2.0.3), $O \subseteq \mathbb{O}$, $a \equiv P_{\mathfrak{M}}(O|i_1)$, and $b \equiv P_{\mathfrak{M}}(O|i_2)$ then we must have $\mathrm{q}(a,b) = T$ and $\mathrm{q}(1-a, 1-b) = T$. The following assumption will help us simplify our subsequent discussion.

ASSUMPTION 2.2.1. *Without loss of generality, we will assume that $\mathrm{q}(a,b) = T \Leftrightarrow \mathrm{q}(1-a, 1-b) = T$ (since we can always replace $\mathrm{q}(a,b)$ with $\mathrm{q}(a,b) \wedge \mathrm{q}(1-a, 1-b)$ without changing the privacy definition).*

Again, we stress that this assumption changes the predicate *without* changing the privacy definition. Our results will thus characterize what $\mathrm{q}(a,b) \wedge \mathrm{q}(1-a, 1-b)$ should look like for a given predicate q to be usable in Definition 2.0.3. The following two results are useful because they will allow us to convert the predicate q into real-valued functions.

PROPOSITION 2.2.2. *Suppose there exists a mechanism $\mathfrak{M}$ that satisfies abstract differential privacy for* q. *Axiom 2.1.1 implies:*

- *$\mathrm{q}(a,a) = T$ for all $a \in [0,1]$.*
- *If $\mathrm{q}(a,b) = T$ and $a \geq 1/2$ then $\mathrm{q}(a,b') = T$ for all b' between b and $\frac{(1-b)a}{1-a}$.*
- *If $\mathrm{q}(a,b) = T$ and $a \leq 1/2$ then $\mathrm{q}(a,b') = T$ for all b' between b and $1 - \frac{b(1-a)}{a}$.*

[2] To make this well-defined, we must assume that $P_{\mathfrak{M}}(\cdot\ |i_1)$ and $P_{\mathfrak{M}}(\cdot\ |i_2)$ have Radon-Nykodin derivatives [36] with respect to the same base measure. For finite and countable output spaces, this condition is vacuous.

PROOF. First note that the existence of a privacy mechanism $\mathfrak{M}$ implies that $\mathrm{q}(1,1) = T$ (and therefore $\mathrm{q}(0,0) = T$) since by Axiom 2.1.1, $\mathcal{A}_1 \circ \mathfrak{M}$ must satisfy privacy whenever $\mathcal{A}_1$ returns the same value o with probability 1 for any input. Now consider $\mathcal{A}_2$ which, on input o outputs o_1 with probability c and o_2 with probability $1-c$. Then $P_{\mathcal{A}_2 \circ \mathcal{A}_1 \circ \mathfrak{M}}(o_1 \mid i) = c$ for any input i. Since $\mathcal{A}_2 \circ \mathcal{A}_1 \circ \mathfrak{M}$ must satisfy abstract differential privacy (by Axiom 2.1.1), we must have $\mathrm{q}(c,c) = T$ for $c \in [0,1]$.

Now create an output space with two points: o_1 and o_2. Define $\mathfrak{M}$ be a randomized algorithm that (1) on input i_1 outputs o_1 with probability a and o_2 with probability $1-a$; and (2) on input i_2 outputs o_1 with probability b and o_2 with probability $1-b$. Clearly $\mathfrak{M}$ is a privacy mechanism for q.

Consider the class of randomized algorithms $\mathcal{A}_{c,d}$ indexed by $c, d \in [0,1]$ such that (1) on input o_1, $\mathcal{A}$ outputs o_1 with probability c and o_2 with probability $1-c$; and (2) on input o_2, $\mathcal{A}$ outputs o_1 with probability d and o_2 with probability $1-d$.

For the case where $a \leq 1/2$, set $d = (1-c)a/(1-a)$. Then as c increases continuously from 0 to 1, d decreases continuously from $a/(1-a)$ to 0. At the same time $P_{\mathcal{A}_{c,d} \circ \mathfrak{M}}(o_1 \mid i_1) = a$ while $P_{\mathcal{A}_{c,d} \circ \mathfrak{M}}(o_1 \mid i_2)$ ranges continuously from $(1-b)a/(1-a)$ to b. Axiom 2.1.1 then implies that $\mathcal{A}_{c,d} \circ \mathfrak{M}$ satisfies abstract differential privacy and so $\mathrm{q}(a,b') = T$ for all b' between b and $(1-b)a/(1-a)$.

For the case where $a \geq 1/2$, we apply our previous result to $1-a$ and $1-b$. Thus $q(1-a, 1-b') = T$ for all $1-b'$ between $1-b$ and $b(1-a)/a$ (and thus all b' between b and $1-b(1-a)/a$). Axiom 2.1.1 then implies that $\mathcal{A}_{c,d} \circ \mathfrak{M}$ satisfies abstract differential privacy and so $q(a,b') = q(1-(1-a), 1-(1-b')) = T$ for all b' between b and $1-b(1-a)/a$. $\square$

The significance of Proposition 2.2.2 is that it allows us to show that for each a, the set of b values that make $\mathrm{q}(a,b) = T$ is actually an interval. We prove this in Proposition 2.2.3.

PROPOSITION 2.2.3. *If there exists a mechanism $\mathfrak{M}$ satisfying abstract differential privacy for* q *then Axiom 2.1.1 implies that there exist functions M_{q} and m_{q} such that for all $a \in [0,1]$, $\mathrm{q}(a,b) = T$ when $m_{\mathrm{q}}(a) < b < M_{\mathrm{q}}(a)$ and $\mathrm{q}(a,b) = F$ whenever $b < m_{\mathrm{q}}(a)$ or $b > M_{\mathrm{q}}(a)$.*

PROOF. By Proposition 2.2.2, for each a there is a b value such that $\mathrm{q}(a,b) = T$. Now, note that if $a \leq 1/2$ and $b \leq a$ then $(1-b)a/(1-a) \geq a$ and if $b \geq a$ then $(1-b)a/(1-a) \leq a$. Similarly, if $a \geq 1/2$ and $b \leq a$ then $1 - \frac{b(1-a)}{a} \geq a$ and if $b \geq a$ then $1 - \frac{b(1-a)}{a} \leq a$.

Fix an $a \in [0,1]$. For each b, Proposition 2.2.2 gives an interval $[low(b), high(b)]$ which contains both b and a such that $\mathrm{q}(a,b') = T$ whenever $b' \in [low(b), high(b)]$. Thus for all b where $\mathrm{q}(a,b) = T$, the corresponding intervals overlap. Thus $\bigcup_{b:\ \mathrm{q}(a,b)=T} [low(b), high(b)]$ is an interval such that $\mathrm{q}(a,b) = T$ if and only if b belongs to this interval. The proof is completed by defining $m_{\mathrm{q}}(a) = \inf \bigcup_{b:\ \mathrm{q}(a,b)=T} [low(b), high(b)]$ and $M_{\mathrm{q}}(a) = \sup \bigcup_{b:\ \mathrm{q}(a,b)=T} [low(b), high(b)]$. $\square$

Thus when the input pair (i_1, i_2) is in our privacy relation $\mathcal{R}$, then given $P_{\mathfrak{M}}(o|i_1)$ there is an interval of allowable values for $P_{\mathfrak{M}}(o|i_2)$. However, the endpoints of the interval may or may not be allowable values. Keeping track of which endpoints are allowable and which are not will greatly complicate the presentation of our ideas, and so we introduce the following proposition which will help simplify things.

PROPOSITION 2.2.4. *Let $\mathfrak{M}$ be a privacy mechanism satisfying abstract differential privacy for* q *and Axiom 2.1.1. Let M_{q} and m_{q} be the functions associated with* q *by Proposition 2.2.3. Let q^* be a predicate such that $\mathrm{q}^*(a,b) = T$ if $b = M_{\mathrm{q}}(a)$ or $b = m_{\mathrm{q}}(a)$ and let $\mathrm{q}^*(a,b) = \mathrm{q}(a,b)$ otherwise. Then $\mathfrak{M}$ is a privacy mechanism for q^* and $M_{\mathrm{q}^*} = M_{\mathrm{q}}$ and $m_{\mathrm{q}^*} = m_{\mathrm{q}}$*

PROOF. The fact that $\mathfrak{M}$ is a privacy mechanism for q^* follows directly from the definition of abstract differential privacy. The rest of the statements follow from the continuity of the $low(b)$ and $high(b)$ functions introduced in the proof of Proposition 2.2.3. $\square$

Thus when studying the properties of m_q and M_q only, we can assume without loss of generality that $\mathrm{q}(a,b) = T$ if and only if $m_{\mathrm{q}} \leq b \leq M_{\mathrm{q}}(a)$. The addition of Axiom 2.1.2 now ensures that the M_q and m_q functions have nice properties.

THEOREM 2.2.5. *For abstract differential privacy (with input space $\mathbb{I}$, output space $\mathbb{O}$, privacy relation $\mathcal{R}$ and set of privacy predicates $\{\mathrm{q}_{i_1,i_2}\}_{(i_1,i_2)\in\mathcal{R}}$), if there exists a privacy mechanism $\mathfrak{M}$ for the privacy predicates $\{\mathrm{q}_{i_1,i_2}\}_{(i_1,i_2)\in\mathcal{R}}$ then*

(i) Axiom 2.1.1 implies that for each $(i_1,i_2) \in \mathcal{R}$ there exist functions M_{i_1,i_2} and m_{i_1,i_2} such that for any $O \subseteq \mathbb{O}$:

$$\begin{aligned} M_{i_1,i_2}(a) > b > m_{i_1,i_2}(a) \quad &\Rightarrow \quad \mathrm{q}_{i_1,i_2}(a,b) = T \\ b > M_{i_1,i_2}(a) \text{ or } b < m_{i_1,i_2}(a) \quad &\Rightarrow \quad \mathrm{q}_{i_1,i_2}(a,b) = F \end{aligned}$$

where $a = P_{\mathfrak{M}}(O \mid i_1)$ and $b = P_{\mathfrak{M}}(O \mid i_2)$.

(ii) Axiom 2.1.1 implies

$$1 \geq M_{i_1,i_2}(a) \geq a \geq m_{i_1,i_2}(a) \geq 0$$

(iii) Axiom 2.1.1 implies

$$M_{i_1,i_2}(a) \geq m_{i_1,i_2}(a) = 1 - M_{i_1,i_2}(1-a)$$

(iv) Axioms 2.1.1 and 2.1.2 imply M_{i_1,i_2} is concave and m_{i_1,i_2} is convex.

(v) Axiom 2.1.1 implies M_{i_1,i_2} is nondecreasing and is strictly increasing at any point a where $M_{i_1,i_2}(a) < 1$. m_{i_1,i_2} is nonincreasing and is strictly decreasing at any point a where $M_{i_1,i_2}(a) > 0$.

(vi) Axioms 2.1.1 and 2.1.2 imply M_{i_1,i_2} is continuous except possibly at $a = 0$ and m_{i_1,i_2} is continuous except possibly at $a = 1$.

PROOF. Fix two points $i_1, i_2 \in \mathbb{I}$ from the input space of $\mathfrak{M}$ such that $(i_1,i_2) \in \mathcal{R}$. To simplify notation we will refer to M_{i_1,i_2} and m_{i_1,i_2} as M and m, respectively. Item (i) is just Proposition 2.2.3. Item (ii) follows easily from the fact that $\mathrm{q}(a,a) = T$ (Proposition 2.2.2). To prove Item (iii), then using Assumption 2.2.1 and Proposition 2.2.4, we have $m(a) \leq b \leq M(a) \Leftrightarrow \mathrm{q}(a,b) = T \Leftrightarrow \mathrm{q}(1-a, 1-b) = T \Leftrightarrow m(1-a) \leq 1-b \leq M(1-a)$ so that $M(a)$ is the maximum allowable value of b if and only if $m(1-a)$ is the minimum allowable value of $1-b$. Item (iii) now follows.

To prove item (iv), consider $a_1 \neq a_2$. Again we invoke Proposition 2.2.4: let $\mathfrak{M}_1$ be the privacy mechanism such that $P_{\mathfrak{M}_1}(o_1 \mid i_1) = a_1$, $P_{\mathfrak{M}_1}(o_2 \mid i_1) = 1-a_1$, $P_{\mathfrak{M}_1}(o_1 \mid i_2) = M(a_1)$ and $P_{\mathfrak{M}_1}(o_2 \mid i_2) = 1 - M(a_1)$. Similarly, let $\mathfrak{M}_2$ be the privacy mechanism such that $P_{\mathfrak{M}_2}(o_1 \mid i_1) = a_2$, $P_{\mathfrak{M}_2}(o_2 \mid i_1) = 1-a_2$, $P_{\mathfrak{M}_2}(o_1 \mid i_2) = M(a_2)$ and $P_{\mathfrak{M}_2}(o_2 \mid i_2) = 1 - M(a_2)$. It is easy to see that $\mathfrak{M}_1$ and $\mathfrak{M}_2$ are privacy mechanisms for q. Now choose a $c \in [0,1]$ and define $\mathfrak{M}_c$ to be the mechanism that runs $\mathfrak{M}_1$ with probability c and $\mathfrak{M}_2$ with probability $1-c$. Axiom 2.1.2 implies that $\mathfrak{M}_c$ is a privacy mechanism for q. Now, $P_{\mathfrak{M}_c}(o_1 \mid i_1) = ca_1 + (1-c)a_2$ and $P_{\mathfrak{M}_c}(o_1 \mid i_2) = cM(a_1) + (1-c)M(a_2)$. Proposition 2.2.3 and the fact that $\mathfrak{M}_c$ is a privacy mechanism for q then implies $M(ca_1 + (1-c)a_2) \geq cM(a_1) + (1-c)M(a_2)$ and so M is concave. The convexity of m then follows from Item (iii).

To prove item (v), choose a such that $M(a) < 1$ and define the mechanism $\mathfrak{M}$ such that $P_{\mathfrak{M}}(o_1 \mid i_1) = a$, $P_{\mathfrak{M}}(o_2 \mid i_1) = 1-a$, $P_{\mathfrak{M}}(o_1 \mid i_2) = M(a)$, and $P_{\mathfrak{M}}(o_2 \mid i_2) = 1-M(a)$ (again we are invoking Proposition 2.2.4). For $0 < c < 1$, define the randomized algorithm $\mathcal{A}_c$ such that $P_{\mathcal{A}_c}(o_1 \mid o_1) = 1$, $P(o_1 \mid o_2) = c$ and $P(o_2 \mid o_2) = 1-c$. Then by Axiom 2.1.1 $\mathcal{A}_c \circ \mathfrak{M}$ is a privacy mechanism for q. Now, $P_{\mathcal{A}_c \circ \mathfrak{M}}(o_1 \mid i_1) = a + c(1-a) > a$ while $M(a + c(1-a)) \geq P_{\mathcal{A}_c \circ \mathfrak{M}}(o_1 \mid i_2) = M(a) + c(1 - M(a)) > M(a)$. Thus M is strictly increasing at any point a where $M(a) < 1$. If $M(a) = 1$ but $a < 1$ then $P_{\mathcal{A}_c \circ \mathfrak{M}}(o_1 \mid i_1) = a + c(1-a) > a$ and $M(a + c(1 - a)) \geq P_{\mathcal{A}_c \circ \mathfrak{M}}(o_1 \mid i_2) = M(a) + c(1 - M(a)) = 1$ and so $M(a + c(1-a)) = 1$ and therefore M is nondecreasing. The corresponding result for m follows from Item (iii).

We now prove Item (vi). Since M is concave (as a result of Axioms 2.1.1 and 2.1.2), a basic continuity result from convexity theory [4] states that M is continuous on the open interval $(0,1)$ (i.e. the relative interior of its domain). Continuity at $a = 1$ follows from the fact that M is nondecreasing and so any discontinuity at 1 would be a jump discontinuity with $M(1) > \epsilon + M(a)$ for some *epsilon* > 0 and all $a < 1$. This contradicts the fact that M is concave. The corresponding result for m follows from Item (iii). □

2.3 Characterizing Abstract Differential Privacy (sufficient conditions)

Here we present the sufficient conditions.

THEOREM 2.3.1. *Let $\mathbb{I}$ be a set and $\mathcal{R} :\subseteq \mathbb{I} \times \mathbb{I} \to \{T, F\}$ and irreflexive predicate. Suppose that for each $(i_1, i_2) \in \mathcal{R}$ there exist functions M_{i_1,i_2} and m_{i_1,i_2} from $[0,1]$ to $[0,1]$ that have the following properties:*

(i) $m_{i_1,i_2}(a) = 1 - M_{i_1,i_2}(1-a)$

(ii) M_{i_1,i_2} *is concave (and* m_{i_1,i_2} *is convex).*

(iii) M_{i_1,i_2} *is continuous on* $(0,1]$ *(and* m_{i_1,i_2} *is continuous on* $[0,1)$*).*

(iv) $M_{i_1,i_2}(0) \geq 0$ *and* $M_{i_1,i_2}(1) = 1$ *(*$m_{i_1,i_2}(0) = 0$ *and* $m_{i_1,i_2}(1) \leq 1$*)*

let $\mathrm{q}_{i_1,i_2}(a,b) = T$ *if and only if* $m_{i_1,i_2}(a) \leq b \leq M_{i_1,i_2}(a)$. *Then Abstract Differential Privacy (Definition 2.0.3) using the predicates* $\{\mathrm{q}_{i_1,i_2}\}_{(i_1,i_2)\in\mathcal{R}}$ *satisfies Axioms 2.1.1 and 2.1.2.*

PROOF. Note that Items (ii) and (iv) and the fact that M_{i_1,i_2} is bounded by 1 ensures that M_{i_1,i_2} is strictly increasing except where it equals 1. Let $\mathfrak{M}$, $\mathfrak{M}_1$, $\mathfrak{M}_2$ be privacy mechanisms for $\{\mathrm{q}_{i_1,i_2}\}_{(i_1,i_2)\in\mathcal{R}}$ with the same input space $\mathbb{I}$ (the existence of such mechanisms is implied by the concavity and nonnegativity of M_{i_1,i_2}, along with $M_{i_1,i_2}(1) = 1$ since then any $\mathfrak{M}$ whose output is independent of the input is a privacy mechanism for $\{\mathrm{q}_{i_1,i_2}\}_{(i_1,i_2)\in\mathcal{R}}$). Fix two points $i_1, i_2 \in \mathbb{I}$ from the input space of $\mathfrak{M}$ such that $(i_1, i_2) \in \mathcal{R}$. To simplify notation we will refer to M_{i_1,i_2} and m_{i_1,i_2} as M and m, respectively.

Implication of Axiom 2.1.1: Transformation Invariance. Choose a randomized algorithm $\mathcal{A}$ (whose input space is the output space of $\mathfrak{M}$) and consider an arbitrary measurable subset S of the output space of $\mathcal{A}$. Let μ_1 be the probability measure $P_{\mathfrak{M}}(\cdot \mid i_1)$ and let μ_2 be the probability measure $P_{\mathfrak{M}}(\cdot \mid i_2)$. Let h_S be the measurable function $P_{\mathcal{A}}(S \mid \cdot)$ and note that $0 \leq h_S \leq 1$. Let $a = P_{\mathcal{A}\circ\mathfrak{M}}(S|i_1)$ and $b = P_{\mathcal{A}\circ\mathfrak{M}}(S|i_2)$. Note that $a = \int h_S(x)\ d\mu_1(x)$ and $b = \int h_s(x)\ d\mu_2(x)$. Our goal is to prove $m(a) \leq b \leq M(a)$. For any measurable subset X of the output space of $\mathfrak{M}$, we will use the notation I_X to denote the indicator function which is 1 on $x \in X$ and 0 otherwise.

Step 1 Suppose $h_S(x) = I_X(x)$ for some measurable subset X of the output space of $\mathfrak{M}$. Then $a = P_{\mathcal{A}\circ\mathfrak{M}}(S|i_1) = \int h_S(x)\ d\mu_1(x) = \mu_1(X) = P_{\mathfrak{M}}(X|i_1)$ and similarly $b = P_{mech}(X|i_2)$ and so since $\mathfrak{M}$ satisfies abstract differential privacy, $m(a) \leq b \leq M(a)$. On the other hand, if $h_S(x) \equiv 0$ then $P_{\mathcal{A}\circ\mathfrak{M}}(S|i_1) = 0$ and $P_{\mathcal{A}\circ\mathfrak{M}}(S|i_2) = 0$. Item (iv) now implies $m(P_{\mathcal{A}\circ\mathfrak{M}}(S \mid i_1)) \leq P_{\mathcal{A}\circ\mathfrak{M}}(S \mid i_2) \leq M(P_{\mathcal{A}\circ\mathfrak{M}}(S \mid i_1))$.

Step 2. We will now prove the theorem for the case when $h_S(x)$ is a *simple function*, that is $h_S(x) = \sum_{j=1}^{n} c_j I_{X_j}(x)$ where the X_j are pairwise disjoint measurable subsets of the output space of $\mathfrak{M}$ and the $c_j \in [0,1]$. Without loss of generality, assume $c_n \leq \cdots \leq c_1$ and for notational convenience, define $c_{n+1} = 0$. In this case,

$$\begin{aligned} a &= P_{\mathcal{A}\circ\mathfrak{M}}(S|i_1) = \int h_S(x)\ d\mu_1(x) = \sum_{j=1}^{n} c_j \mu_1(X_j) \\ b &= P_{\mathcal{A}\circ\mathfrak{M}}(S|i_2) = \int h_S(x)\ d\mu_2(x) = \sum_{j=1}^{n} c_j \mu_2(X_j) \end{aligned}$$

we can rewrite a and b as follows:

$$a = c_1 \sum_{j=1}^{n} \frac{c_j - c_{j+1}}{c_1} \mu_1\left(\bigcup_{\ell=1}^{j} X_\ell\right) \quad (1)$$

$$b = c_1 \sum_{j=1}^{n} \frac{c_j - c_{j+1}}{c_1} \mu_2\left(\bigcup_{\ell=1}^{j} X_\ell\right) \quad (2)$$

and note that the factors $\frac{c_j - c_{j+1}}{c_1}$ are nonnegative, sum up to 1, and therefore define a convex combination. From Step 1 we have for all j:

$$m\left(\mu_1\left(\bigcup_{\ell=1}^{j} X_\ell\right)\right) \leq \mu_2\left(\left(\bigcup_{\ell=1}^{j} X_\ell\right)\right) \leq M\left(\mu_1\left(\bigcup_{\ell=1}^{j} X_\ell\right)\right)$$

Thus

$$\begin{aligned} m(a) &= m\left(c_1 \sum_{j=1}^{n} \frac{c_j - c_{j+1}}{c_1} \mu_1\left(\bigcup_{\ell=1}^{j} X_\ell\right)\right) \\ &\leq c_1 m\left(\sum_{j=1}^{n} \frac{c_j - c_{j+1}}{c_1} \mu_1\left(\bigcup_{\ell=1}^{j} X_\ell\right)\right) \\ &\quad \text{(since } m \text{ is convex and } m(0) = 0\text{)} \\ &\leq c_1 \sum_{j=1}^{n} \frac{c_j - c_{j+1}}{c_1} m\left(\mu_1\left(\bigcup_{\ell=1}^{j} X_\ell\right)\right) \\ &\quad \text{(by convexity of } m\text{)} \\ &\leq c_1 \sum_{j=1}^{n} \frac{c_j - c_{j+1}}{c_1} \mu_2\left(\bigcup_{\ell=1}^{j} X_\ell\right) \\ &\quad \text{by Step 1} \\ &= b \quad \text{(by Equation 2)} \end{aligned}$$

$$\begin{aligned} M(a) &= M\left(c_1 \sum_{j=1}^{n} \frac{c_j - c_{j+1}}{c_1} \mu_1\left(\bigcup_{\ell=1}^{j} X_\ell\right)\right) \\ &\geq c_1 M\left(\sum_{j=1}^{n} \frac{c_j - c_{j+1}}{c_1} \mu_1\left(\bigcup_{\ell=1}^{j} X_\ell\right)\right) \\ &\quad \text{(since } M \text{ is concave and } M(0) \geq 0\text{)} \\ &\geq c_1 \sum_{j=1}^{n} \frac{c_j - c_{j+1}}{c_1} M\left(\mu_1\left(\bigcup_{\ell=1}^{j} X_\ell\right)\right) \\ &\geq c_1 \sum_{j=1}^{n} \frac{c_j - c_{j+1}}{c_1} \mu_2\left(\bigcup_{\ell=1}^{j} X_\ell\right) \\ &= b \end{aligned}$$

Step 3. We now prove the theorem for arbitrary measurable $h_S(x)$. By Theorem 1.17 in [36], there exists a sequence $h_S^{(1)}, h_S^{(2)}, \ldots$ of simple functions such that for all x, $0 \leq h_S^{(1)}(x) \leq h_S^{(2)}(x) \leq \cdots \leq h_S(x)$ and $\lim_{n\to\infty} h_S^{(n)}(x) \to h_S(x)$. By the Lebesgue Monotone Convergence Theorem [36],

$$\begin{aligned} \lim_{n\to\infty} \int h_S^{(n)}(x)\, d\mu_1(x) &\to \int h_S(x)\, d\mu_1(x) = P_{\mathcal{A}\circ\mathfrak{M}}(S|i_1) \\ \lim_{n\to\infty} \int h_S^{(n)}(x)\, d\mu_2(x) &\to \int h_S(x)\, d\mu_2(x) = P_{\mathcal{A}\circ\mathfrak{M}}(S|i_2) \end{aligned}$$

From Step 2 we have: $m\left(\int h_S^{(n)}(x)\, d\mu_1(x)\right) \leq \int h_S^{(n)}(x)\, d\mu_2(x) \leq M\left(\int h_S^{(n)}(x)\, d\mu_1(x)\right)$ The continuity of M (except at 0) then implies that $M\left(\int h_S(x)\, d\mu_1(x)\right) \geq \int h_S(x)\, d\mu_2(x)$ except possibly in the case when $\int h_S(x)\, d\mu_1(x) = 0$. However, $\int h_S(x)\, d\mu_1(x) = 0$ implies that $h_S(x) \equiv 0$ except possibly on a set X with $\mu_1(X) = 0$ (since $h_S(x)$ cannot be negative); this case is covered by Step 1.

Similarly, the continuity of m (except at 1) then implies that $m\left(\int h_S(x)\, d\mu_1(x)\right) \leq \int h_S(x)\, d\mu_2(x)$ except possibly in the case when $\int h_S(x)\, d\mu_1(x) = 1$. However, since $h_S(x) \leq 1$ then $\int h_S(x)\, d\mu_1(x) = 1$ implies that $h_S(x) = I_X(x)$ for some measurable set X and so this case is also covered by Step 1.

Implication of Axiom 2.1.2: Privacy Axiom of Choice. Now consider privacy mechanisms $\mathfrak{M}_1$ and $\mathfrak{M}_2$ with the same input space. Choose $c \in [0, 1]$ and defined $\mathfrak{M}_c$ as the randomized algorithm that on input i it returns $\mathfrak{M}_1(i)$ with probability c and $\mathfrak{M}_2(i)$ with probability $1 - c$ (independently of the input). Let S be an arbitrary measurable subset of the union of the output spaces of $\mathfrak{M}_1$ and $\mathfrak{M}_2$. Thus $m(P_{\mathfrak{M}_1}(S \mid i_1)) \leq P_{\mathfrak{M}_1}(S \mid i_2) \leq M(P_{\mathfrak{M}_1}(S \mid i_1))$ and $m(P_{\mathfrak{M}_2}(S \mid i_1)) \leq P_{\mathfrak{M}_2}(S \mid i_2) \leq M(P_{\mathfrak{M}_2}(S \mid i_1))$. Now, $P_{\mathfrak{M}_p}(S \mid i_1) = pP_{\mathfrak{M}_1}(S \mid i_1) + (1-p)P_{\mathfrak{M}_2}(S \mid i_1)$ and $P_{\mathfrak{M}_p}(S \mid i_2) = pP_{\mathfrak{M}_1}(S \mid i_2) + (1-p)P_{\mathfrak{M}_2}(S \mid i_2)$. By the convexity of m and concavity of M, we have

$$\begin{aligned} m(P_{\mathfrak{M}_p}(S \mid i_1)) &= m(pP_{\mathfrak{M}_1}(S \mid i_1) + (1-p)P_{\mathfrak{M}_2}(S \mid i_1)) \\ &\leq pm(P_{\mathfrak{M}_1}(S \mid i_1)) + (1-p)m(P_{\mathfrak{M}_2}(S \mid i_1)) \\ &\leq pP_{\mathfrak{M}_1}(S \mid i_2) + (1-p)P_{\mathfrak{M}_2}(S \mid i_2) \\ &= P_{\mathfrak{M}_p}(S \mid i_2) \\ M(P_{\mathfrak{M}_p}(S \mid i_1)) &\geq pM(P_{\mathfrak{M}_1}(S \mid i_1)) + (1-p)M(P_{\mathfrak{M}_2}(S \mid i_1)) \\ &\geq pP_{\mathfrak{M}_1}(S \mid i_2) + (1-p)P_{\mathfrak{M}_2}(S \mid i_2) \\ &= P_{\mathfrak{M}_p}(S \mid i_2) \end{aligned}$$

□

3. REASONING ABOUT UTILITY

To really take advantage of privacy definitions (both new and old), we need to design privacy mechanisms that output the most useful data possible. For example, any mechanism whose output is independent of the input satisfies generic differential privacy. However, this is not a pleasing result since it seems that we can do "better". A common approach for "doing better" is to select a utility measure arbitrarily or with the justification that it is used in decision-theoretic statistics.

Although intuitively this seems like a valid approach, recent results indicate otherwise. Ghosh et al. [23] have shown, in the case of differential privacy, that if a user asks a single count query, believes in a prior distribution over query answers, and provides a loss function from a suitably well-behaved class then the following is true. There exists a privacy mechanism, called the *geometric mechanism* [23], such that any optimal mechanism (in the sense of minimizing expected loss) can be constructed from the geometric mechanism by a lossy postprocessing step (in general, the geometric mechanism is not optimal by this utility metric). This postprocessing step is a deterministic function that is not one-to-one and thus removes information.

Thus, expected utility seems like a poor choice of utility metric when choosing a privacy mechanism. In addition, optimizing a privacy mechanism $\mathfrak{M}$ for one specific task may also be a mistake – there could exist a privacy mechanism $\mathfrak{M}'$ such that $\mathfrak{M}(i) = \mathcal{A}(\mathfrak{M}'(i))$ for some randomized algorithm $\mathcal{A}$. Thus choosing $\mathfrak{M}'$ instead of the highly tuned $\mathfrak{M}$ would be preferable because $\mathfrak{M}'$ is clearly just as useful for the original task, but may also be useful for other tasks as well. In this sense $\mathfrak{M}'$ is more *general*.

We axiomatically formalize this notion of generality in Section 3.1. We then present several measures of utility and discuss whether or not they are appropriate for use in statistical privacy (Section 3.2). Finally, we characterize what optimal differentially private mechanisms should look like for finite input and output spaces and then specialize

this result to a utility measure known as the negative of Dobrushin's coefficient of ergodicity [11] (Section 3.3).

3.1 The Generality Axiom for Utility

Recall that a privacy mechanism $\mathfrak{M}$ is a randomized algorithm with input space $\mathbb{I}$ and output space $\mathbb{O}$ satisfying some privacy definition. We represent $\mathfrak{M}$ as a conditional probability distribution $P_{\mathfrak{M}}(o \mid i)$ just as with any randomized algorithm. For any randomized algorithm $\mathcal{A}$, $\mathcal{A} \circ \mathfrak{M}$ denotes the composition of $\mathcal{A}$ and $\mathfrak{M}$ defined by first running $\mathfrak{M}$ and then running $\mathcal{A}$ on the output of $\mathfrak{M}$.

When the input and output spaces are finite we treat $\mathfrak{M}$ as a column stochastic matrix[3] $\{m_{j,k}\}$ whose (j,k) entry $m_{j,k}$ is equal to $P(j|k)$. Thus the rows correspond to elements of the output space and columns correspond to elements of the input space. We will abuse notation and use the symbol $\mathfrak{M}$ to refer to the matrix as well. In matrix form, the composition $\mathcal{A} \circ \mathfrak{M}$ is equivalent to $\mathcal{A}\,\mathfrak{M}$ (interpreted as matrix multiplication).

Convention 3.1.1. (Matrix form of $\mathcal{A}$) *Given a randomized algorithm $\mathcal{A}$ with finite input and output space, we represent $\mathcal{A}$ as a matrix $\{m_{i,j}\}$ such that $m_{i,j} = P_{\mathcal{A}}(i \mid j)$.*

We will also need to define a partial order on the set of privacy mechanisms that satisfy a given privacy definition:

Definition 3.1.2. (Generality Partial Order). *Let S be the set of privacy mechanisms that satisfy a particular privacy definition. If $\mathfrak{M}_1 \in S$ and $\mathfrak{M}_2 \in S$ then we say that $\mathfrak{M}_2$ is at least as general as $\mathfrak{M}_1$, and denote this by $\mathfrak{M}_1 \preceq_{\mathrm{G}} \mathfrak{M}_2$, if there exists a randomized algorithm $\mathcal{A}$ such that the conditional probability distribution $P_{\mathfrak{M}_1}$ is equal to $P_{\mathcal{A} \circ \mathfrak{M}_2}$. We call this partial order the* generality partial order.

Thus if you can probabilistically simulate $\mathfrak{M}_1$ by postprocessing the output of $\mathfrak{M}_2$ with some randomized algorithm $\mathcal{A}$, then $\mathfrak{M}_2$ is considered at least as general as $\mathfrak{M}_1$. It would also appear to be at least as preferable for this reason.

Definition 3.1.3. (Maximally General). *Let S be the set of privacy mechanisms that satisfy a particular privacy definition. A privacy mechanism $\mathfrak{M} \in S$ is* maximally general *if for every privacy-mechanism $\mathfrak{M}' \in S$ such that $\mathfrak{M} \preceq_{\mathrm{G}} \mathfrak{M}'$ it is also true that $\mathfrak{M}' \preceq_{\mathrm{G}} \mathfrak{M}$.*

Give a set S of privacy mechanisms that satisfy a given privacy definition, the subset S_{Max} of maximally general mechanisms is clearly desirable, especially if it is non-empty and has a certain coverage property: for every $\mathfrak{M} \in S$ there exists a $\mathfrak{M}^* \in S_{Max}$ such that $\mathfrak{M} \preceq_{\mathrm{G}} \mathfrak{M}^*$ (that is, every privacy mechanism can be realized as the postprocessing of the output of a maximally general mechanism). In general, for arbitrary privacy definitions satisfying privacy Axioms 2.1.1 and 2.1.2, S_{Max} will not be guaranteed to have these nice properties. Of course, we could add an axiom to force S_{Max} to have these properties, but as of now we are unable to justify it in terms of privacy alone (since it would be a privacy axiom). Thus we did not include such an axiom in Section 2.1. Nevertheless maximal generality is a useful concept and leads naturally to the following axiom:

[3]A matrix with nonnegative entries where each column sums up to 1.

Axiom 3.1.4. (Generality Axiom). *A measure μ of the utility of a privacy mechanism must respect the generality partial order $\preceq_{\mathrm{G}}$. That is, $\mu(\mathfrak{M}) \geq \mu(\mathcal{A} \circ \mathfrak{M})$ for any randomized algorithm whose input space is the output space of $\mathfrak{M}$.*

3.2 Measures of Generality

In this section, using Axiom 3.1.4, we examine some candidate measures of utility. For simplicity, we will assume that the input and output spaces are finite and thus we treat privacy mechanisms and randomized algorithms as column stochastic matrices, as discussed in Convention 3.1.1. Note that we need our utility measures μ to satisfy the following property: $\mu(\mathcal{A}\,\mathfrak{M}) \leq \mu(\mathfrak{M})$ (where $\mathfrak{M}$ is a privacy mechanism and $\mathcal{A}$ is a randomized algorithm).

Example 3.2.1. (Negative Expected Loss). *Let L be a loss matrix where $L(j,k)$ is the loss we incur for outputting j when k is the true input. If $\mathfrak{M}$ is a privacy mechanism, we may want to minimize its expected loss, which is equivalent to maximizing $-Trace(L^T\,\mathfrak{M})$. The results of Ghosh et al. [23] imply that negative expected utility does not satisfy Axiom 3.1.4.*

Example 3.2.2. (Absolute value of Determinant). *If $\mathfrak{M} = \{m_{i,j}\}$ is represented as a square column stochastic matrix (see Convention 3.1.1) then it seems natural to consider the utility measure $\mu(\mathfrak{M}) = |det(\mathfrak{M})|$. The multiplicative properties of the determinant show that $|det(\mathcal{A}\,\mathfrak{M})| = |det(\mathcal{A})|\,|\det(\mathfrak{M})| \leq |\det(\mathfrak{M})|$ for a randomized algorithm $\mathcal{A}$ that is represented by a square matrix, since column stochastic matrices have determinants with absolute value ≤ 1. Geometrically, this measures the contractive properties of $\mathfrak{M}$ because $\mathfrak{M}$ maps the unit hypercube into another convex polytope whose area is $|det(\mathfrak{M})|$ [36]. This μ satisfies our utility criterion with the proviso that we are only considering privacy mechanisms whose output space is the same as the input space. This is a restrictive assumption since we show in Section 3.3 that for differential privacy there are many maximally general privacy mechanisms with much larger output spaces.*

Example 3.2.3. *For a privacy mechanism $\mathfrak{M} = \{m_{i,j}\}$, define $\mu_{Dob}(\mathfrak{M}) = -\min_{j,k} \sum_i \min(m_{i,j}, m_{i,k})$. This is the negative of Dobrushin's coefficient of ergodicity and is another useful measure of the contractive properties (in the geometric sense) of a column stochastic matrix [11]. We prove that $\mu_{Dob}(\mathcal{A}\,\mathfrak{M}) \leq \mu_{Dob}(\mathfrak{M})$ in Appendix A and we characterize optimal differentially private mechanisms (in terms of μ_{Dob}) in Section 3.3.*

3.3 Characterizing the Dobrushin Coefficient of Ergodicity

In this section we characterize "optimal" differentially private mechanisms. Our main results are Theorem 3.3.3, which characterizes what maximally general mechanisms with finite input and output spaces look like, and Lemma 3.3.8, which characterizes optimal differentially private mechanisms (with finite input and output spaces) according to μ_{Dob}, the negative Dobrushin coefficient of ergodicity (see Example 3.2.3).

Before presenting the technical parts of these results, we first informally discuss our results and their consequences. Recall that according to our view of differential privacy,

there is an input space $\mathbb{I}$ and output space $\mathbb{O}$, a *symmetric* privacy relation $\mathcal{R} \subseteq \mathbb{I} \times \mathbb{I}$, and the constraint that for all measurable $O \subseteq \mathbb{O}$ and $(i_1, i_2) \in \mathcal{R}$ we must have $P_{\mathfrak{M}}(O \mid i_1) \leq e^{\epsilon} P_{\mathfrak{M}}(O \mid i_2)$ or $P_{\mathfrak{M}}(O \mid i_2) \leq e^{\epsilon} P_{\mathfrak{M}}(O \mid i_1)$.

For each $o \in \mathbb{O}$, if we look at all edges in $\mathcal{R}$ where either of those constraints holds with equality, we would get a subgraph for this output o (formally, we call this a row graph; see Definition 3.3.2). Theorem 3.3.3 states that for a maximally general mechanism $\mathfrak{M}$, a necessary and sufficient condition is that the subgraph for each o contains a spanning tree of the privacy relation $\mathcal{R}$. By identifying o with its subgraph, we see that the natural output space is the set of spanning trees of $\mathcal{R}$ – it is not a set of query answers or possible datasets.

What is the interpretation of such an output space? Noting that the subgraph represents constraints that hold with equality and that the subgraph spans $\mathcal{R}$, it is easy to see that for a given $o \in \mathbb{O}$, the associated subgraph uniquely determines (up to multiplicative constant) the likelihood function $L_o(i)$ defined as $L_o(i) = P_{\mathfrak{M}}(o|i)$. In other words, the output space should be a restricted subset of likelihood functions. To statisticians this is a pleasing result, since according to the *likelihood principle* [9], the likelihood function is all we need for statistical inference.

On the other hand, this result is less pleasing to end-users, who use statistical software whose input is data, not likelihood functions. Thus, in addition to maximally general mechanisms, we need to develop additional tools to shoehorn this output space into a format that can be digested by off-the-shelf statistical software.

Our second result characterizes optimal mechanisms under the negative Dobrushin coefficient in Lemma 3.3.8. The essence of the lemma is that there must exist a graph structure that is common to the subgraphs of all $o \in \mathbb{O}$. The only thing that differs is the pattern of which constraints ($P_{\mathfrak{M}}(o \mid i_1) \leq e^{\epsilon} P_{\mathfrak{M}}(o \mid i_2)$ or $P_{\mathfrak{M}}(o \mid i_1) \geq e^{-\epsilon} P_{\mathfrak{M}}(o \mid i_2)$) are tight. We note here that the Geometric Mechanism [23] (also defined in Definition 3.3.1) satisfies the conditions of the lemma and thus is optimal under the negative Dobrushin coefficient.

DEFINITION 3.3.1. (Geometric Mechanism [23][4]). *Let* $\mathbb{I} = 1, \ldots, n$, *let* $\mathbb{O} = 1, \ldots, n$, *and let* $\mathcal{R} = \{(1,2), (2,3), \ldots, (n-1, n)\}$. *Given* $\epsilon > 0$, *the geometric mechanism is a randomized algorithm that on input* i *selects an integer* z *with probability proportional to* $e^{-\epsilon|i-z|}$. *If* $z > n$ *then it resets* z *to* n. *If* $z < 1$ *it resets* z *to 1.*

Our proof shows that many mechanisms can maximize the negative Dobrushin coefficient, including ones that are not maximally general. However, we can apply the constructive process we use in the proof of Theorem 3.3.3 to convert them to maximally general mechanisms (without increasing or decreasing the value of the Dobrushin coefficient). This leads to an interesting philosophical question: should we require an axiom stronger than Axiom 3.1.4 which requires that only maximally general mechanisms can maximize a utility measure? In theory we do not gain anything because under Axiom 3.1.4, if the mechanism we derived from an optimization process is not general, we can always make it more general without affecting the utility score. However, on an intuitive level a stronger axiom does make more sense. For now we leave this issue unresolved.

The technical part of this section begins with the concept of the *row-graphs* of a differentially private mechanism $\mathfrak{M}$, which mark the places where the constraints enforced by differential privacy are true with equality.

DEFINITION 3.3.2. (Row graphs). *For a differentially private mechanism* $\mathfrak{M}$ *with finite output space, the* row graphs *of* $\mathfrak{M}$ *are a set of graphs, one for each* $o \in \mathbb{O}$. *The graph associated with output* o *has* $\mathbb{I}$ *as the set of nodes, and for any* $i_1, i_2 \in \mathbb{I}$, *there is an edge* (i_i, i_2) *if* $(i_1, i_2) \in \mathcal{R}$ *and either* $P_{\mathfrak{M}}(O \mid i_1) = e^{\epsilon} P_{\mathfrak{M}}(O \mid i_2)$ *or* $P_{\mathfrak{M}}(O \mid i_2) = e^{\epsilon} P_{\mathfrak{M}}(O \mid i_1)$.

The following theorem formally shows that some common intuition - trying to maximize the number of equality constraints in differential privacy - is a consequence of Axiom 3.1.4. The theorem is also instructive in the sense that it shows that the output space can be much bigger than the input space (an upper bound on its size is the number of spanning trees of $\mathcal{R}$ when viewed as a graph).

Previous work on differential privacy has focused on mechanisms with output spaces that were at most the size of the input space or were equivalent (according to the $\preceq_{\mathrm{G}}$ partial order) to such mechanisms. Many maximally general mechanisms could have been missed this way. Thus the existence of parts of this theorem are obvious after the fact, but surprisingly not *a priori*. The proof is in Appendix B.

THEOREM 3.3.3. *For a given* $\epsilon > 0$, *finite input space* $\mathbb{I}$, *and privacy relation* $\mathcal{R}$ *(it must be symmetric for differential privacy), let* S_{con} *be the set of all differentially private mechanisms* $\mathfrak{M}$ *with finite output spaces and such that each row graph is a connected graph. Then* S_{con} *is precisely the set of maximally general differentially private mechanisms with finite output spaces.*

Recall that for finite input and output spaces, we treat a privacy mechanism $\mathfrak{M}$ as a column stochastic matrix $\{m_{i,j}\}$ corresponding to its conditional probabilities (see Convention 3.1.1). For convenience, we define the following function:

$$E_{j,k}(\mathfrak{M}) = -\sum_i \min(m_{i,j}, m_{i,k})$$

and note that $\mu_{Dob}(\mathfrak{M}) = \max_{j,k} E_{j,k}(\mathfrak{M})$. We also define the linear privacy relation

$$\mathcal{R}_n = \{(i, i+1) \ : \ i = 1, \ldots, n-1\}$$

where we have identified any input space $\mathbb{I}$ of size n with the first n positive integers. It is instructive to study the privacy relations $\mathcal{R}_n$ first as our results for general[5] $\mathcal{R}$ depend on $\mathcal{R}_n$ via reduction.

The following lemmas consider mechanisms $\mathfrak{M} = \{m_{i,j}\}$ represented as $2 \times n$ matrices with privacy relations $\mathcal{R}_n$. These results will help us in the general case.

LEMMA 3.3.4. *Let* $\epsilon > 0$ *and let* $\mathfrak{M} = \{m_{i,j}\}$ *be a differentially private mechanism that is represented as a* $2 \times n$ *matrix with privacy relation* $\mathcal{R}_n$. *If* $\mathfrak{M}$ *maximizes the function* $E_{1,n}$ *in the class of* $2 \times n$ *mechanisms then either* $m_{1,j} = \max(e^{-\epsilon} m_{1,j+1}, 1 - (1 - m_{j+1})e^{\epsilon})$ *for* $j = 1, \ldots, n-1$ *or* $m_{1,j+1} = \max(e^{-\epsilon} m_{1,j}, 1 - (1 - m_j)e^{\epsilon})$ *for* $j = 1, \ldots, n-1$.

[4]Note that this formulation is equivalent to the formulation presented by Ghosh et al. [23]

[5]i.e. any symmetric privacy relation $\mathcal{R}$ that is connected when viewed as a graph.

The proof is in Appendix E

LEMMA 3.3.5. *Let $\epsilon > 0$ and let $\mathfrak{M} = \{m_{i,j}\}$ be a differentially private mechanism that is represented as a $2 \times n$ matrix with privacy relation $\mathcal{R}_n$. $\mathfrak{M}$ maximizes the function $E_{1,n}$ in the class of $2 \times n$ mechanisms (and without loss of generality $m_{1,1} \leq m_{1,n}$) if and only if*

- *if n is even $m_{1,j} = e^{\epsilon} m_{1,j-1}$ for $j = 2, \ldots, n/2+1$ and $m_{2,j} = e^{-\epsilon} m_{2,j-1}$ for $j = n/2+1, \ldots, n$.*
- *if n is odd then $m_{1,j} = e^{\epsilon} m_{1,j-1}$ for $j = 2, \ldots, (n+1)/2$ and $m_{2,j} = e^{-\epsilon} m_{2,j-1}$ for $j = (n+3)/2, \ldots, n$.*

The proof is included in Appendix C.

DEFINITION 3.3.6. (Distance function $d_{\mathcal{R}}(i,j)$). *Given input $i, j \in \mathbb{I}$ and a symmetric relation $\mathcal{R}$ that is connected when viewed as a graph the function $d_{\mathcal{R}}(i,j)$ is the length of the shortest path from i to j in $\mathcal{R}$.*

LEMMA 3.3.7. *Let $\mathbb{I}$ be a finite input space with $|\mathbb{I}| > 1$ and let $\mathcal{R}$ be a symmetric, connected relation. Let d be the diameter of $\mathcal{R}$ when viewed as a graph (i.e. distance between furthest two nodes). If $\mathfrak{M} = \{m_{i,j}\}$ is a differentially private mechanism that maximizes $\mu_{Dob}(\mathfrak{M}) = \max_{j,k} E_{j,k}(\mathfrak{M})$, then for any two $j, k \in \mathbb{I}$ for which $\mu_{Dob}(\mathfrak{M}) = E_{j,k}(\mathfrak{M})$ we must have $d_{\mathcal{R}}(j,k) = d$.*

PROOF. Without loss of generality, assume $\mathfrak{M}$ has no 0 entries (otherwise an entire row would contain only 0 and we can safely remove it).

Suppose there exist $u_1, v_1 \in \mathbb{I}$ such that $\mu_{Dob}(\mathfrak{M}) = E_{u_1,v_1}(\mathfrak{M})$ and $d_{\mathcal{R}}(u_1, v_1) = t < d$. We will create a $\mathfrak{M}^* = \{m^*_{i,j}\}$ with $\mu_{Dob}(\mathfrak{M}^*) > \mu_{Dob}(\mathfrak{M})$, thus causing a contradiction.

Choose a u_2, v_2 with $d_{\mathcal{R}}(u_2, v_2) = t + 1$. We partition $\mathbb{I}$ into groups $G_1, \ldots, G_{t+1}$ based on distance from u_2. For $k \leq t$, a node $i \in G_k$ if $d_{\mathcal{R}}(u_2, i) = k$, and $i \in G_{t+1}$ if $d_{\mathcal{R}}(u_2, i) \geq t + 1$. $\mathfrak{M}^*$ will be constant within groups: if i_1 and i_2 are in the same group then $P_{\mathfrak{M}^*}(\cdot \mid i_1) = P_{\mathfrak{M}^*}(\cdot \mid i_2)$.

Let $i_0 = u_1$, $i_t = v_1$, and let $i_0, i_1, \ldots, i_t$ be the nodes on a shortest path from u_1 to v_1. For any $o \in \mathbb{O}$, and $k \leq t$, set

$$P_{\mathfrak{M}^*}(o \mid j) = P_{\mathfrak{M}}(o \mid i_k) \quad \text{for } j \in G_k$$

We now deal with G_{t+1}. There must exist $o_1, o_2 \in \mathbb{O}$ such that $P_{\mathfrak{M}}(o_1 \mid i_0) \leq P_{\mathfrak{M}}(o_1 \mid i_t)$ and $P_{\mathfrak{M}}(o_2 \mid i_0) \geq P_{\mathfrak{M}}(o_2 \mid i_t)$ (otherwise the probabilities could not sum to 1). Choose $\delta > 0$ such that

$$\frac{P_{\mathfrak{M}}(o_1 \mid i_t) + \delta}{P_{\mathfrak{M}}(o_1 \mid i_t)} \leq e^{\epsilon} \quad \textit{and} \quad \frac{P_{\mathfrak{M}}(o_2 \mid i_t) - \delta}{P_{\mathfrak{M}}(o_2 \mid i_t)} \geq e^{-\epsilon}$$

Then for $j \in G_{t+1}$, set $P_{\mathfrak{M}^*}(o_1 \mid j) = P_{\mathfrak{M}}(o_1 \mid i_t) + \delta$, $P_{\mathfrak{M}^*}(o_2 \mid j) = P_{\mathfrak{M}}(o_2 \mid i_t) - \delta$, and $P_{\mathfrak{M}^*}(o \mid j) = P_{\mathfrak{M}}(o \mid i_t)$ for all other $o \in \mathbb{O}$.

Clearly $\mu_{Dob}(\mathfrak{M}^*) = \mu_{Dob}(\mathfrak{M}) + \delta > \mu_{Dob}(\mathfrak{M})$ and so we just need to check the differential privacy constraints. For any input $u \in \mathbb{I}$ assigned to group G_k, its neighbors are either in the same group G_k or in a group G_{k-1} or G_{k+1}. Thus by construction we ensured that the differential privacy constraints continue to hold. □

LEMMA 3.3.8. *Let $\mathbb{I}$ be a finite input space with $|\mathbb{I}| = n > 1$ and let $\mathcal{R}$ be a symmetric, connected relation. Let d be the diameter of $\mathcal{R}$. If $\mathfrak{M} = \{m_{i,j}\}$ is a differentially private mechanism maximizes μ_{Dob} if and only if there exist $u, v \in \mathbb{I}$ such that $d_{\mathcal{R}}(u,v) = d$ and $\mu_{Dob}(\mathfrak{M}) = E_{u,v}(\mathfrak{M})$ and some shortest path $u = i_0, i_1, \ldots, i_d = v$ such that for all $r \in \mathbb{O}$:*

- *if d is even, then either $m_{r,i_j} = e^{\epsilon} m_{r,i_{j-1}}$ for $j = 2, \ldots, d/2+1$ or $m_{r,i_j} = e^{-\epsilon} m_{r,i_{j-1}}$ for $j = d/2+1, \ldots, d$.*
- *if d is odd then $m_{r,i_j} = e^{\epsilon} m_{r,i_{j-1}}$ for $j = 2, \ldots, (d+1)/2$ and $m_{r,i_j} = e^{-\epsilon} m_{r,i_{j-1}}$ for $j = (d+3)/2, \ldots, d$.*

PROOF. For necessary conditions, first note that Lemma 3.3.7 implies $d_{\mathcal{R}}(u,v) = d$ is necessary and $\mu_{Dob}(\mathfrak{M}) = E_{u,v}(\mathfrak{M})$. Define the randomized algorithm $\mathcal{A}$ that outputs either 1 or 0 such that for any $o \in \mathbb{O}$, $P_{\mathcal{A}}(0 \mid o) = 1$ if $P_{\mathfrak{M}}(o \mid u) \leq P_{\mathfrak{M}}(o \mid v)$ and $P_{\mathcal{A}}(1 \mid o) = 1$ if $P_{\mathfrak{M}}(o \mid u) > P_{\mathfrak{M}}(o \mid v)$. Note that $\mu_{Dob}(\mathfrak{M}) = \mu_{Dob}(\mathcal{A} \circ \mathfrak{M})$ because for those $o \in \mathbb{O}$ that were mapped into 1, $P_{\mathfrak{M}}(o \mid u)$ contributes to $E_{u,v}(\mathfrak{M})$ while $P_{\mathfrak{M}}(o \mid v)$ contributes in all other cases. If we restrict $\mathcal{A} \circ \mathfrak{M}$ to the inputs $i_0, \ldots, _d$, then we have a $2 \times n$ mechanism that must maximize the function $E_{1,d}$. Lemma 3.3.5 then implies the necessary conditions. It also implies sufficient conditions since the application of $\mathcal{A}$ would not change μ_{Dob}. □

4. RELATED WORK

Our efforts at axiomatizing privacy and utility are motivated by corresponding efforts in mathematical philosophy and probabilistic inductive logic (e.g., [8, 33]) where the goal is to model the reasoning of a rational agent.

For surveys on statistical privacy, see [10, 21, 1].

The need for a better understanding of privacy definitions and privacy mechanisms was underscored by the work of Dinur and Nissim [12] (and later by Dwork et al. [17]) that showed that some intuitive methods for preserving privacy actually did not preserve privacy according to essentially any privacy definition. This work was followed by a line of research that led to differential privacy [6, 14, 16, 15, 32, 13]. Note that there have been some attempts to weaken the definition of differential privacy (e.g., [15, 32, 29, 27]) as its stringent guarantees are not always considered necessary (especially when data utility can be increased).

What sets differential privacy apart from most privacy definitions is the strength of its guarantees and the ability to formally investigate its properties. In particular, Rastogi et al. provide a connection between privacy and utility guarantees [35] as well as a connection to another definition known as adversarial privacy [34], which was also studied by Evfimievski et al. [19] in the context of query auditing.

Utility of sanitized data has also been studied. Of particular relevance are the following. McSherry and Talwar [30] have presented a general recipe for taking a "quality function" and turning it into a privacy mechanism for differential privacy. Although this recipe does not come with guarantees, it has been used successfully in other work [7]. Dwork et al. [18] provided a link between utility and computational complexity for differential privacy. Recent work by Ghosh et al. [23] has shown that optimizing for commonly accepted utility metrics (in this case expected utility) is not always the correct goal since the output of a suboptimal mechanism (according to the utility metric) may sometimes be post-processed (in a lossy way) to mimic the output of an "optimal" mechanism.

5. CONCLUSIONS

In this paper we presented three axioms for privacy and utility and showed how they can guide the development of

privacy definitions, utility measures, and privacy mechanisms. We feel this is just the beginning of a unified theory of privacy. Additional consequences of the generality axiom of utility need to be explored, especially in the context of generic differential privacy. We also plan to explore axioms concerning prior beliefs that an attacker may possess. Currently many privacy definitions have not been expressed formally enough to apply an axiomatic approach. We feel this makes them into privacy goals rather privacy definitions and additional work is needed to formalize them so that they can be analyzed under the same mathematical lens as differential privacy.

6. ACKNOWLEDGMENTS

We would like to thank Adam Smith from Penn State University for helpful discussions.

7. REFERENCES

[1] Nabil Adam and John Wortmann. Security-control methods for statistical databases. *ACM Computing Surveys*, 21(4):515–556, 1989.

[2] B. Barak, K. Chaudhuri, C. Dwork, S. Kale, F. McSherry, and K. Talwar. Privacy, accuracy, and consistency too: A holistic solution to contingency table release. In *PODS*, 2007.

[3] Michael Barbaro and Tom Zeller. A face is exposed for AOL searcher no. 4417749. *New York Times*, August 9 2006.

[4] Dimitri P. Bertsekas, Angelia Nedic, and Asuman E. Ozdaglar. *Convex Analysis and Optimization*. Athena Scientific, 2003.

[5] U. Blien, H. Wirth, and M. Muller. Disclosure risk for microdata stemming from official statistics. *Statistica Neerlandica*, 46(1):69–82, 1992.

[6] Avrim Blum, Cynthia Dwork, Frank McSherry, and Kobbi Nissim. Practical privacy: the sulq framework. In *PODS*, pages 128–138, 2005.

[7] Avrim Blum, Katrina Ligett, and Aaron Roth. A learning theory approach to non-interactive database privacy. In *STOC*, pages 609–618, 2008.

[8] Rudolf Carnap and Richard C. Jeffrey, editors. *Studies in Inductive Logic and Probability*, volume I. University of California Press, 1971.

[9] George Casella and Roger L. Berger. *Statistical Inference*. Duxbury, 2nd edition, 2002.

[10] Bee-Chung Chen, Daniel Kifer, Kristen LeFevre, and Ashwin Machanavajjhala. Privacy-preserving data publishing. *Foundations and Trends in Databases*, 2(1-2):1–167, 2009.

[11] Joel E. Cohen, Yves Derriennic, and Gh. Zbaganu. Majorization, monotonicity of relative entropy and stochastic matrices. *Contemporary Mathematics*, 149, 1993.

[12] Irit Dinur and Kobbi Nissim. Revealing information while preserving privacy. In *PODS*, 2003.

[13] C. Dwork and N. Nissim. Privacy-preserving datamining on vertically partitioned databases. In *CRYPTO*, 2004.

[14] Cynthia Dwork. Differential privacy. In *ICALP*, volume 4051 of *Lecture Notes in Computer Science*, pages 1–12, 2006.

[15] Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. Our data, ourselves: Privacy via distributed noise generation. In *EUROCRYPT*, pages 486–503, 2006.

[16] Cynthia Dwork, Frank Mcsherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography Conference*, pages 265–284, 2006.

[17] Cynthia Dwork, Frank McSherry, and Kunal Talwar. The price of privacy and the limits of lp decoding. In *STOC*, pages 85–94, 2007.

[18] Cynthia Dwork, Moni Naor, Omer Reingold, Guy N.Rothblum, and Salil Vadhan. On the complexity of differentially private data release: Efficient algorithms and hardness results. In *STOC*, pages 381–390, 2009.

[19] Alexandre Evfimievski, Ronald Fagin, and David P. Woodruff. Epistemic privacy. In *PODS*, 2008.

[20] Alexandre Evfimievski, Johannes Gehrke, and Ramakrishnan Srikant. Limiting privacy breaches in privacy-preserving data mining. In *PODS*, 2003.

[21] B. Fung, K. Wang, R. Chen, and P. Yu. Privacy-preserving data publishing: A survey on recent developments. *ACM Computing Surveys*, 42(4), 2010.

[22] Srivatsava Ranjit Ganta, Shiva Prasad Kasiviswanathan, and Adam Smith. Composition attacks and auxiliary information in data privacy. In *KDD*, 2008.

[23] Arpita Ghosh, Tim Roughgarden, and Mukund Sundararajan. Universally utility-maximizing privacy mechanisms. In *STOC*, pages 351–360, 2009.

[24] M. Hay, V. Rastogi, G. Miklau, and D. Suciu. Boosting the accuracy of differentially-private histograms through consistency. In *VLDB*, 2010.

[25] Daniel Kifer. Attacks on privacy and de finetti's theorem. In *SIGMOD*, 2009.

[26] Ravi Kumar, Jasmine Novak, Bo Pang, and Andrew Tomkins. On anonymizing query logs via token-based hashing. In *WWW*, 2007.

[27] Ashwin Machanavajjhala, Johannes Gehrke, and Michaela Götz. Data publishing against realistic adversaries. *VLDB*, 2009.

[28] Ashwin Machanavajjhala, Johannes Gehrke, Daniel Kifer, and Muthuramakrishnan Venkitasubramaniam. ℓ-diversity: Privacy beyond k-anonymity. In *ICDE*, 2006.

[29] Ashwin Machanavajjhala, Daniel Kifer, John Abowd, Johannes Gehrke, and Lars Vilhuber. Privacy: Theory meets practice on the map. *ICDE*, pages 277–286, 2008.

[30] Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *FOCS*, pages 94–103, 2007.

[31] M. Ercan Nergiz and Chris Clifton. Thoughts on k-anonymization. *Data & Knowledge Engineering*, 63(3):622–645, 2007.

[32] Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. Smooth sensitivity and sampling in private data analysis. In *STOC*, pages 75–84, 2007.

[33] C. J. Nix and J. B. Paris. A continuum of inductive methods arising from a generalized principle of instantial relevance. *Journal of Philosophical Logic*,

35(1):83–115, 2006.

[34] Vibhor Rastogi, Michael Hay, Gerome Miklau, and Dan Suciu. Relationship privacy: Output perturbation for queries with joins. In *PODS*, pages 107–116, 2009.

[35] Vibhor Rastogi, Dan Suciu, and Sungho Hong. The boundary between privacy and utility in data publishing. In *VLDB*, pages 531–542, 2007.

[36] Walter Rudin. *Real & Complex Analysis.* McGraw-Hill, 3rd edition, 1987.

[37] Pierangela Samarati and Latanya Sweeney. Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. Technical report, CMU, SRI, 1998.

[38] Mark J. Schervish. *Theory of Statistics.* Springer, 1995.

[39] Latanya Sweeney. k-anonymity: a model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10(5):557–570, 2002.

[40] Raymond Wong, Ada Fu, Ke Wang, and Jian Pei. Minimality attack in privacy preserving data publishing. In *VLDB*, 2007.

[41] Xiaokui Xiao, Guozhang Wang, and Johannes Gehrke. Differential privacy via wavelet transforms. In *ICDE*, 2010.

APPENDIX

A. DOBRUSHIN'S COEFFICIENT OF ERGODICITY AND MAXIMAL GENERALITY

Here we prove that the negative Dobrushin coefficient of ergodicity, here defined as $\mu_{Dob}(\mathfrak{M}) = -\min_{j,k}\sum_i \min(m_{i,j}, m_{i,k})$, satisfies the relation $\mu_{Dob}(\mathcal{A}\,\mathfrak{M}) \leq \mu_{Dob}(\mathfrak{M})$.

PROOF. Let $\mathfrak{M}$ be a privacy mechanism with column-stochastic matrix representation $\{m_{i,j}\}$. Let $\mathcal{A}$ be a randomized algorithm with column-stochastic matrix representation $\{p_{i,j}\}$, with appropriate dimensions so that the product $\mathcal{A}\,\mathfrak{M}$ makes sense.

Below, we will use the fact that min is concave and $c\min(x_1, x_2) = \min(cx_1, cx_2)$ for $c \geq 0$ from which it follows that $\min(\sum_{i=1}^r p_i x_i) \geq \sum_{i=1}^r p_i \min(x_i)$ when $p_i \geq 0$ for all i.

$$\begin{aligned}
\sum_i \min(m_{i,j}, m_{i,k}) &= \sum_i \sum_\ell p_{\ell,i} \min\left(m_{i,j}, m_{i,k}\right) \\
&= \sum_\ell \sum_i p_{\ell,i} \min\left(m_{i,j}, m_{i,k}\right) \\
&\leq \sum_\ell \min\left(\sum_i m_{i,j} p_{\ell,i}, \sum_i m_{i,k} p_{\ell,i}\right) \\
&= \sum_\ell \min\left(m'_{\ell,j}, m'_{\ell,k}\right)
\end{aligned}$$

where $\{m'_{\ell,j}\}$ is the matrix representation of $\mathcal{A}\,\mathfrak{M}$. Thus it follows that $\min_{j,k}\sum_i \min(m_{i,j}, m_{i,k}) \leq \min_{j,k}\sum_\ell \min\left(m'_{\ell,j}, m'_{\ell,k}\right)$ and so $\mu_{Dob}(\mathfrak{M}) \geq \mu_{Dob}(\mathcal{A}\,\mathfrak{M})$.

□

B. MAXIMALLY GENERAL DIFFERENTIALLY PRIVATE MECHANISMS WITH FINITE INPUT SPACES

In this section we characterize the maximally general differentially private mechanisms for the case when the input space is finite. Recall that according to our view of differential privacy with finite input space, there is a finite input space $\mathbb{I}$ (of sensitive data), and output space $\mathbb{O}$, a *symmetric* privacy relation $\mathcal{R} \subseteq \mathbb{I} \times \mathbb{I}$, and the constraint that for all measurable $O \subseteq \mathbb{O}$ and $(i_1, i_2) \in \mathcal{R}$ we must have $P_{\mathfrak{M}}(O \mid i_1) \leq e^\epsilon P_{\mathfrak{M}}(O \mid i_2)$.

We first need to introduce the concept of the *row-graphs* of a differentially private mechanism $\mathfrak{M}$, which mark the places where the constraints enforced by differential privacy are true with equality.

DEFINITION B.0.9. (Row graphs). *For a differentially private mechanism $\mathfrak{M}$ with finite output space, the* row graphs *of $\mathfrak{M}$ are a set of graphs, one for each $o \in \mathbb{O}$. The graph associated with row o has $\mathbb{I}$ as the set of nodes, and for any $i_1, i_2 \in \mathbb{I}$, there is an edge (i_i, i_2) if $(i_1, i_2) \in \mathcal{R}$ and either $P_{\mathfrak{M}}(O \mid i_1) = e^\epsilon P_{\mathfrak{M}}(O \mid i_2)$ or $P_{\mathfrak{M}}(O \mid i_2) = e^\epsilon P_{\mathfrak{M}}(O \mid i_1)$.*

The following theorem formally shows that some common intuition - trying to maximize the number of equality constraints in differential privacy - is a consequence of Axiom 3.1.4. The theorem is also instructive in the sense that it shows that the output space can be much bigger than the input space (an upper bound on its size is the number of spanning trees of $\mathcal{R}$ when viewed as a graph).

Previous work on differential privacy has focused on mechanisms with output spaces that were at most the size of the input space or were equivalent (according to the $\preceq_{\mathrm{G}}$ partial order) to such mechanisms. Many maximally general mechanisms could have been missed this way. Thus the existence of parts of this theorem are obvious after the fact, but surprisingly not *a priori*.

THEOREM B.0.10. *For a given $\epsilon > 0$, finite input space $\mathbb{I}$, and privacy relation $\mathcal{R}$ (it must be symmetric for differential privacy), let S be the set of all differentially private mechanisms. Let S_{con} be the subset of S consisting of all mechanisms $\mathfrak{M}$ with finite output spaces and such that each row graph is connected. Then S_{con} is precisely the set of maximally general differentially private mechanisms with finite output spaces.*

PROOF. When viewed as a column stochastic matrix, no maximally general mechanism can have an entry equal to 1 (the constraints for differential privacy would then imply that an entire row consists of entries equal to 1, meaning that the output of such a mechanism is constant). Such a mechanism is clearly not in S_{con} (the row containing all 1 entries is not connected).

We first show that mechanisms with finite output spaces excluded from S_{con} cannot be maximally general. Let $\mathfrak{M}$ be a mechanism and let o be an output such that the corresponding row graph is not connected. This row can be decomposed into two disjoint components C_1 and C_2 such that there are no edges between them. Let

$$\rho_1 = \max\{e^\delta : \exists s \in C_1, t \in C_2, P(o|s) = e^\delta P(o|t)\} \quad (3)$$

$$\rho_2 = \max\{e^\delta : \exists s \in C_1, t \in C_2, P(o|t) = e^\delta P(o|s)\} \quad (4)$$

and note that $0 < \rho_1 < e^\epsilon$ and $0 < \rho_2 < e^\epsilon$ (if either ρ_1 or ρ_2 were 0 then the whole row of $\mathfrak{M}$ would consist entirely of 0's and all constraints would be tight). Define

$$\begin{aligned} a &= \frac{(e^\epsilon/\rho_2) - 1}{(e^\epsilon/\rho_1)(e^\epsilon/\rho_2) - 1} \\ b &= \frac{(e^\epsilon/\rho_1) - 1}{(e^\epsilon/\rho_1)(e^\epsilon/\rho_2) - 1} \end{aligned}$$

We form a new output space $\mathbb{O}' = \mathbb{O} \setminus \{o\} \uplus \{o_1, o_2\}$ (where $\uplus$ denotes disjoint union) by splitting o into two outputs o_1 and o_2 and define mechanism $\mathfrak{M}'$ with output space $\mathbb{O}'$ such that

$$P_{\mathfrak{M}'}(o' \mid s) = \begin{cases} P_{\mathfrak{M}}(o|s) \times ae^\epsilon/\rho_1 & \text{if } o' = o_1 \ \wedge \ s \in C_1 \\ P_{\mathfrak{M}}(o|s) \times a & \text{if } o' = o_1 \ \wedge \ s \in C_2 \\ P_{\mathfrak{M}}(o|s) \times b & \text{if } o' = o_2 \ \wedge \ s \in C_1 \\ P_{\mathfrak{M}}(o|s) \times be^\epsilon/\rho_2 & \text{if } o' = o_2 \ \wedge \ s \in C_2 \\ P_{\mathfrak{M}}(o' \mid s) & \text{if } o' \in \mathbb{O} \cap \mathbb{O}' \end{cases}$$

Note that all of these are proper probabilities since a, b, ae^ϵ/ρ_1, be^ϵ/ρ_2 are nonnegative and less than 1 since $e^\epsilon > \rho_1$ and $e^\epsilon > \rho_2$. Clearly we also must have for each fixed s, $\sum_{o' \in \mathbb{O}} P_{\mathfrak{M}'}(o' \mid s) = 1$.

Let $\mathcal{A}$ be a randomized algorithm such that $\mathcal{A}(o') = o'$ if $o' \in \mathbb{O} \cap \mathbb{O}'$ and $\mathcal{A}(o') = o$ if $o' \in \{o_1, o_2\}$. We claim that:

- $\mathfrak{M} \preceq_{\mathrm{G}} \mathfrak{M}'$. Proof: clearly $\mathfrak{M} = \mathcal{A} \circ \mathfrak{M}'$.
- $\mathfrak{M}'$ satisfies differential privacy: if $s, t \in C_1$ then clearly $\frac{P_{\mathfrak{M}'}(o_i|s)}{P_{\mathfrak{M}'}(o_i|t)} = \frac{P_{\mathfrak{M}}(o|s)}{P_{\mathfrak{M}}(o|t)}$ for $i = 1, 2$ (and similarly for $s, t \in C_2$). If $s \in C_1$ and $t \in C_2$ then since the row corresponding to o cannot have 0 entries:

  $$\frac{P_{\mathfrak{M}'}(o_1|s)}{P_{\mathfrak{M}'}(o_1|t)} = P_{\mathfrak{M}}(o|s)/P_{\mathfrak{M}}(o|t) \times e^\epsilon/\rho_1 \leq e^\epsilon \quad (5)$$

  Since $P_{\mathfrak{M}}(o|s)/P_{\mathfrak{M}}(o|t) \leq \rho_1$. We reach a similar conclusion for $s \in C_2$ and $t \in C_1$. The results for o_2 are similar.
- The row graph of o with respect to $\mathfrak{M}$ is a proper subgraph of the row graphs of o_1 and o_2 with respect to $\mathfrak{M}'$: since there is no edge between C_1 and C_2 in the row graph of o with respect to $\mathfrak{M}$, then the previous argument shows that any edge present in the row graph for o is also present in the row graphs for o_1 and o_2. Also note that in Equation 5, equality is achieved for the s and t that achieve the maximum in Equation 3. Thus the row graph for o_1 has an additional edge. Similarly, the row graph for o_2 has an additional edge.
- The randomized algorithm $\mathcal{A}$ defined above is not reversible so $\mathfrak{M}'$ is strictly more general than $\mathfrak{M}$ (the proof is obvious).

Repeating this procedure finitely many times (the number is at most the number of spanning trees in the privacy relation $\mathcal{R}$ when viewed as a graph), we get a privacy mechanism that belongs to S_{con}. Thus there can be no maximally general differentially private mechanism with finite output space that does not belong to S_{con}.

To show that every $\mathfrak{M} \in S_{con}$ is maximally general, first note that if any two rows of $\mathfrak{M}$ are proportional to each other (which can only happen if the corresponding row graphs are the same), we can form a mechanism $\mathfrak{M}_2$ which is the same as $\mathfrak{M}$ except that those two rows are replaced by one row containing their sum. It is easy to see that $\mathfrak{M} \preceq_{\mathrm{G}} \mathfrak{M}'$ and $\mathfrak{M}' \preceq_{\mathrm{G}} \mathfrak{M}$. Thus for this part of the proof it is enough to assume that now two rows of $\mathfrak{M}$ are proportional to each other and no two row graphs are the same.

Now , suppose there exists a privacy mechanism $\mathfrak{M}'$ with output space $\mathbb{O}'$ and a randomized algorithm $\mathcal{A}$ such that $P_{\mathcal{A} \circ \mathfrak{M}'} = P_{\mathfrak{M}}$ for some $\mathfrak{M} \in S_{con}$ with output space $\mathbb{O}$. For any $o \in \mathbb{O}$, define $\mathcal{A}^-(o) \equiv \{o' \in \mathbb{O} \ : \ P_{\mathcal{A}}(o \mid o') > 0\}$ (it is a poor man's inverse). It is easy to see that every measurable $O' \subseteq \mathcal{A}^-(o)$, and any $(i_1, i_2) \in \mathcal{R}$, $\frac{P_{\mathfrak{M}}(o|i_1)}{P_{\mathfrak{M}}(o|i_2)} = \frac{P_{\mathfrak{M}'}(O'|i_1)}{P_{\mathfrak{M}'}(O'|i_2)}$ whenever the denominator of the right hand side is nonzero (in which case the numerator must also be positive by the differential privacy requirements). This is because the tightness constraints in the row graph for o determine (up to a constant factor) all the probabilities $P_{\mathfrak{M}}(o \mid \cdot)$ and no positive combination of nontight constraints can yield a tight constraint for differential privacy. This implies that the conditional probability $P_{\mathfrak{M}'}(O' \mid \mathcal{A}^-(o), i)$ is independent of the input i.

Since no other row in $\mathfrak{M}$ has the same row graph as o (without loss of generality) and all other row graphs are connected and therefore each represent a maximal set of tight constraints in differential privacy, we see that $P_{\mathcal{A}}(o \mid o') = 1$ for all $o' \in \mathcal{A}^-(o)$. This implies that $P_{\mathfrak{M}}(o \mid i) = P_{\mathfrak{M}'}(\mathcal{A}^-(o) \mid i)$ for any $i \in \mathbb{I}$.

Thus we can define a randomized algorithm $\mathcal{A}_2$ that for any $o \in \mathbb{O}$ and $O' \subseteq \mathcal{A}^-(o)$, we have $P_{\mathcal{A}_2}(O' \mid o) = P_{\mathfrak{M}'}(O' \mid \mathcal{A}^-(o), i)$ (for any $i \in \mathbb{I}$ since this quantity does not depend on i). Using the fact that $P_{\mathfrak{M}}(o \mid i) = P_{\mathfrak{M}'}(\mathcal{A}^-(o) \mid i)$ for any $i \in \mathbb{I}$, it is each to check that $P_{\mathfrak{M}'} = P_{\mathcal{A}_2 \circ \mathfrak{M}}$ and so $\mathfrak{M}' \preceq_{\mathrm{G}} \mathfrak{M}$. Thus we have shown that every $\mathfrak{M} \in S_{con}$ is maximally general.

□

C. CHARACTERIZING DOBRUSHIN'S COEFFICIENT

Here we restate and prove Lemma 3.3.5.

LEMMA C.0.11. *Let $\epsilon > 0$ and let $\mathfrak{M} = \{m_{i,j}\}$ be a differentially private mechanism that is represented as a $2 \times n$ matrix with privacy relation $\mathcal{R}_n$. $\mathfrak{M}$ maximizes the function $E_{1,n}$ in the class of $2 \times n$ mechanisms (and without loss of generality $m_{1,1} \leq m_{1,n}$) if and only if*

- *if n is even $m_{1,j} = e^\epsilon m_{1,j-1}$ for $j = 2, \ldots, n/2+1$ and $m_{2,j} = e^{-\epsilon} m_{2,j-1}$ for $j = n/2+1, \ldots, n$.*
- *if n is odd then $m_{1,j} = e^\epsilon m_{1,j-1}$ for $j = 2, \ldots, (n+1)/2$ and $m_{2,j} = e^{-\epsilon} m_{2,j-1}$ for $j = (n+3)/2, \ldots, n$.*

PROOF. First we prove necessary conditions. Note that Lemma 3.3.4 holds. For a mechanism $\mathfrak{M} = \{m_{i,j}\}$ that is represented as a $2 \times n$ matrix, $m_{1,j} = \max(e^{-\epsilon} m_{1,j+1}, 1 - (1 - m_{j+1})e^\epsilon)$ implies that $m_{1,j+1} = \min(e^\epsilon m_{1,j}, 1 - e^{-\epsilon}(1 - m_{1,j}))$. Furthermore simple calculations show that $e^\epsilon m_{1,j} \leq 1 - e^{-\epsilon}(1 - m_{1,j})$ if and only if $m_{1,j} \leq \frac{1}{1+e^\epsilon}$ and this is true if and only if $m_{1,j+1} \leq \frac{e^\epsilon}{1+e^\epsilon}$. Combined with the results from Lemma 3.3.4, we have

$$m_{1,j} = e^\epsilon m_{1,j-1} \quad \text{iff} \quad m_{1,j} \leq \frac{e^\epsilon}{1+e^\epsilon} \quad (6)$$

$$m_{2,j} = e^{-\epsilon} m_{2,j-1} \quad \text{iff} \quad m_{2,j} \leq \frac{1}{1+e^\epsilon} \quad (7)$$

Now suppose there is no index ℓ such that $\frac{1}{1+e^\epsilon} < m_{1,\ell} \leq \frac{e^\epsilon}{1+\epsilon}$. Then $m_{1,n} \leq 1/(1+e^\epsilon)$ or $m_{1,1} > e^\epsilon/(1+e^\epsilon)$. If $m_{1,n} \leq 1/(1+e^\epsilon)$ then we must have $m_{1,1}e^{\epsilon(n-1)} = m_{1,n}$ and

$$\begin{aligned} E_{1,n}(\mathfrak{M}) &= -m_{1,1} - m_{2,n} \\ &= -m_{1,1} - (1 - m_{1,1}e^{\epsilon(n-1)}) \end{aligned}$$

and this is an increasing function of $m_{1,1}$ so we should increase $m_{1,1}$ until $m_{1,n} = e^\epsilon/(1+e^\epsilon)$, which is the largest value $m_{1,n}$ can take for which the relation $m_{1,1}e^{\epsilon(n-1)} = m_{1,n}$ still holds (thus we arrive at a contradiction). A similar argument holds for the case where $m_{1,1} > e^\epsilon/(1+e^\epsilon)$. Since $m_{1,n} \geq m_{1,1}$, these arguments have shown something even more important: no matter what, we must have $m_{1,n} \geq e^\epsilon/(1+e^\epsilon)$ (we will use this soon).

Thus there is an index ℓ such that $\frac{1}{1+e^\epsilon} < m_{1,\ell} \leq \frac{e^\epsilon}{1+\epsilon}$ which then implies $m_{1,j} = e^\epsilon m_{1,j-1}$ for $j = 2, \ldots, \ell$ and $m_{2,j} = e^{-\epsilon}m_{2,j-1}$ for $j = \ell+1, \ldots, n$. Thus $m_{1,1} = e^{-\epsilon(\ell-1)}m_{1,\ell}$ and $m_{2,n} = (1 - m_{1,\ell})e^{-\epsilon(n-\ell)}$ and therefore

$$E_{1,n}(\mathfrak{M}) = -e^{-\epsilon(\ell-1)}m_{1,\ell} \quad - \quad (1 - m_{1,\ell})e^{-\epsilon(n-\ell)} \tag{8}$$

We have two parameters to optimize – the constrained value of $m_{1,\ell}$ and the integer ℓ.

First, suppose n is even. Then $\ell - 1 \neq n - \ell$. Equation 8 is a linear function of $m_{1,\ell}$ and (for fixed ℓ) is maximized when $m_{1,\ell}$ is at the appropriate endpoint $\frac{1}{1+e^\epsilon}$ or $\frac{e^\epsilon}{1+\epsilon}$. If the best value is $1/(1+e^\epsilon)$ then $m_{1,\ell+1} = e^\epsilon/(1+e^\epsilon)$ (remember $\ell+1$ exists since we have shown that $m_{1,n} \geq \frac{e^\epsilon}{1+e^\epsilon}$), and thus we could not have chosen the index ℓ since $\ell + 1$ satisfies the conditions we placed when choosing an index. Thus $m_{1,\ell} = \frac{e^\epsilon}{1+e^\epsilon}$. Equation 8 then becomes:

$$\begin{aligned} E_{1,n}(\mathfrak{M}) &= -e^{-\epsilon(\ell-1)}\frac{e^\epsilon}{1+e^\epsilon} \quad - \quad \frac{1}{1+e^\epsilon}e^{-\epsilon(n-\ell)} \\ &= -e^{-\epsilon(\ell-2)}\frac{1}{1+e^\epsilon} \quad - \quad \frac{1}{1+e^\epsilon}e^{-\epsilon(n-\ell)} \end{aligned} \tag{9}$$

This is clearly a concave function of ℓ and it is easy to see (by setting derivatives to 0) that it is maximized when $\ell - 2 = n - \ell$ so $\ell = n/2 + 1$. Note that we must also have $m_{2,\ell} = \frac{1}{1-e^\epsilon}$. Applying Equations 6 and 7, we have the case for even n.

Now suppose n is odd. If $\ell - 1 \neq n - \ell$ (i.e. $\ell \neq (n+1)/2$) then a similar argument leads to Equation 9 and $m_{1,\ell} = e^\epsilon/(1+e^\epsilon)$. We would like to set $\ell = n/2 + 1$ however, ℓ must be an integer, and so using the fact that Equation 9 is concave, ℓ must be one of the nearest integers, either $(n+1)/2$ or $(n+3)/2$. Since we supposed $\ell \neq (n+1)/2$, we must use $\ell = (n+3)/2$. From Equation 9, we have $E_{1,n} = -[e^{-\epsilon(n-1)/2} + e^{-\epsilon(n-3)/2}]/(1+e^\epsilon) = -e^{-\epsilon(n-1)/2}$.

Now, in the other case where $\ell - 1 = n - \ell$ (i.e. $\ell = (n+1)/2$), Equation 8 simplifies to $-e^{-\epsilon(n-1)/2}$ and $m_{1,\ell}$ can be any number in $(\frac{1}{1+e^\epsilon}, \frac{e^\epsilon}{1+e^\epsilon}]$. So whether $\ell = (n+1)/2$ or $\ell = (n+3)/2$, Equations 6 and 7 then imply the Lemma for odd n.

For sufficient conditions, it is clear that for even n, the mechanism is determined uniquely. For odd n, there are many choices but all yield the same value. □

D. SEMANTIC INTERPRETATION OF GENERIC DIFFERENTIAL PRIVACY

Here we restate and prove Proposition 2.1.4

Suppose i_1 is the true data. An attacker may have a prior belief in the probability of i_1 and i_2. We express this as the log-odds $\log(\frac{P_{attacker}(i_2)}{P_{attacker}(i_1)})$. If $\mathfrak{M}(i_1)$ outputs some $o \in \mathbb{O}$ then the attacker's log odds will become $\log(\frac{P_{attacker}(i_2 \mid o)}{P_{attacker}(i_1 \mid o)})$. Denote the difference between them as $\Delta = \log(\frac{P_{attacker}(i_2 \mid o)}{P_{attacker}(i_1 \mid o)}) - \log(\frac{P_{attacker}(i_2)}{P_{attacker}(i_1)})$. The probability that Δ takes a value x is then the probability that any bad $o \in \mathbb{O}$ is produced which changes the log-odds by x. This random variable Δ has the following behavior:

PROPOSITION D.0.12. *Let i_1 be the true data and let $\mathfrak{M}$ be a privacy mechanism for generic differential privacy.*[6] *If $P_{attacker}(i_1) > 0$ and $P_{attacker}(i_2) > 0$ then for $\epsilon > 0$ we have $P(\Delta \geq \epsilon \mid i_1) \leq a'$ where $a' = \sup\{a > 0 : \log\frac{M_{i_1,i_2}(a)}{a} \geq \epsilon\}$. Similarly, $P(\Delta \leq -\epsilon \mid i_1) \leq a''$ where $a'' = \sup\{a > 0 : \log\frac{m_{i_1,i_2}(a)}{a} \leq -\epsilon\}$ (with the convention that $\sup\emptyset = 0$). In both cases the probability depends only on the randomness in $\mathfrak{M}$.*

PROOF. We first need to show that $\log\frac{M_{i_1,i_2}(a)}{a} \geq 0$ and is nonincreasing for $a \in (0,1]$ while $\log\frac{m_{i_1,i_2}(a')}{a'} \leq 0$ and is a nondecreasing function of $a \in (0,1]$. Theorem 2.2.5, Item (ii) shows that the first function is nonnegative and the second is nonpositive. Since M_{i_1,i_2} is nonnegative and concave, then for $c \in (0,1)$ we have $\frac{M_{i_1,i_2}(ca)}{ca} \geq c\frac{M_{i_1,i_2}(a)}{ca} + (1-c)\frac{M_{i_1,i_2}(0)}{ca} \geq \frac{M_{i_1,i_2}(a)}{a}$. Similarly, by definition, m_{i_1,i_2} is convex and $m_{i_1,i_2}(0) = 0$. Thus for $c \in (0,1)$, $\frac{m_{i_1,i_2}(ca)}{ca} \leq c\frac{m_{i_1,i_2}(a)}{ca} + (1-c)\frac{m_{i_1,i_2}(0)}{ca} = \frac{m_{i_1,i_2}(a)}{a}$. The rest follows from the monotonicity of log.

Now, using Bayes' Theorem, we get

$$\begin{aligned} \Delta &= \log\frac{P_{attacker}(i_2)P_{\mathfrak{M}}(o \mid i_2)}{P_{attacker}(i_1)P_{\mathfrak{M}}(o \mid i_1)} - \log\frac{P_{attacker}(i_2)}{P_{attacker}(i_1)} \\ &= \log\frac{P_{\mathfrak{M}}(o \mid i_2)}{P_{\mathfrak{M}}(o \mid i_1)} \end{aligned}$$

Consider the set $O_{bad} = \{o \in \mathbb{O} \mid \frac{P_{\mathfrak{M}}(o \mid i_2)}{P_{\mathfrak{M}}(o \mid i_1)} \geq e^\epsilon\}$. Clearly $\frac{M_{i_1,i_2}(P_{\mathfrak{M}}(O_{bad} \mid i_1))}{P_{\mathfrak{M}}(O_{bad} \mid i_1)} \geq \frac{P_{\mathfrak{M}}(O_{bad} \mid i_2)}{P_{\mathfrak{M}}(O_{bad} \mid i_1)} \geq e^\epsilon$ and since $\frac{M_{i_1,i_2}(a)}{a}$ is nonincreasing, $P(\Delta \geq \epsilon \mid i_1) \leq P(O_{bad} \mid i_1) \leq a'$. A similar argument yields the corresponding result for $P(\Delta \leq -\epsilon \mid i_1)$. □

E. RESULTS FOR $2 \times N$ MECHANISMS

In this section we restate and prove Lemma 3.3.4.

LEMMA E.0.13. *Let $\epsilon > 0$ and let $\mathfrak{M} = \{m_{i,j}\}$ be a differentially private mechanism that is represented as a $2 \times n$ matrix with privacy relation $\mathcal{R}_n$. If $\mathfrak{M}$ maximizes the function $E_{1,n}$ in the class of $2 \times n$ mechanisms then either $m_{1,j} = \max(e^{-\epsilon}m_{1,j+1}, 1 - (1 - m_{j+1})e^\epsilon)$ for $j = 1, \ldots, n-1$ or $m_{1,j+1} = \max(e^{-\epsilon}m_{1,j}, 1 - (1 - m_j)e^\epsilon)$ for $j = 1, \ldots, n-1$.*

PROOF. Without loss of generality, assume $m_{1,1} \leq m_{1,n}$ and note that $m_{2,1} \geq m_{2,n}$ because each column must sum to 1 (and we only have two rows). It is easy to see that neither row can contain a 0 for $\mathfrak{M}$ that maximize $E_{1,n}$.

[6] To make this well-defined, we must assume that $P_{\mathfrak{M}}(\cdot \mid i_1)$ and $P_{\mathfrak{M}}(\cdot \mid i_2)$ have Radon-Nykodin derivatives [36] with respect to the same base measure. For finite and countable output spaces, this condition is vacuous.

Notice that for a given value of $m_{1,j+1}$, the quantity $\max(e^{-\epsilon}m_{1,j+1}, 1-(1-m_{j+1})e^{\epsilon})$ is the smallest value that $m_{1,j}$ can take while still satisfying the differential privacy conditions: $e^{\epsilon}m_{1,j+1} \geq m_{1,j} \geq e^{-\epsilon}m_{1,j+1}$ and $e^{-\epsilon}m_{2,j+1} \leq m_{2,j} \leq e^{\epsilon}m_{2,j+1}$.

So, by way of contradiction, suppose that $\mathfrak{M} = \{m_{i,j}\}$ maximizes $E_{1,n}$ but violates the condition that $m_{1,\ell} = \max(e^{-\epsilon}m_{1,\ell+1}, 1-(1-m_{\ell+1})e^{\epsilon})$ for some ℓ. Since $\mathfrak{M}$ is differentially private, we must have $m_{1,\ell} > \max(e^{-\epsilon}m_{1,\ell+1}, 1-(1-m_{\ell+1})e^{\epsilon})$.

Choose the smallest ℓ for which this is true. We construct a mechanism $\mathfrak{M}^* = \{m^*_{i,j}\}$ with $E_{1,n}(\mathfrak{M}^*) > E_{1,n}(\mathfrak{M})$ as follows. Inductively set

$$\begin{aligned} m^*_{1,j} &= m_{1,j} \quad \text{for } \ell+1 \leq j \leq n \\ m^*_{1,j} &= \max(e^{-\epsilon}m^*_{1,j+1}, 1-(1-m^*_{j+1})e^{\epsilon}) \quad \text{for } 1 \leq j \leq \ell \\ m^*_{2,j} &= 1-m^*_{1,j} \quad \text{for } 1 \leq j \leq n \end{aligned}$$

It is easy to check that $\mathfrak{M}^*$ satisfies differential privacy. Furthermore, given the value of $m^*_{i,j+1}$ (for $j = 1, \ldots, \ell$), $m^*_{1,j}$ is as small as possible. Combined with the fact that $m^*_{1,\ell} < m_{1,\ell}$ (by construction), we must have $m^*_{1,j} < m_{1,j}$ for $j \leq \ell$ and $m^*_{1,j} = m_{1,j}$ for $j > \ell$.

Since we only have two rows (so $m_{1,n} = 1 - m_{2,n}$), and by assumption $m_{1,1} \leq m_{1,n}$ (see first statement of the proof), $E_{1,n}(\mathfrak{M}) = -(m_{1,1} + 1 - m_{1,n})$ and $E_{1,n}(\mathfrak{M}^*) = -(m^*_{1,1} + 1 - m^*_{1,n})$. Using the preceding facts, we get $E_{1,n}(\mathfrak{M}) < E_{1,n}(\mathfrak{M}^*)$ contradicting the maximality of $E_{1,n}(\mathfrak{M})$. $\square$