Journal of
Cybersecurity and Privacy

MDPI

Article

Refinement Orders for Quantitative Information Flow and Differential Privacy †

Konstantinos Chatzikokolakis [1], **Natasha Fernandes** [2,3] **and Catuscia Palamidessi** [3,*]

[1] Department of Informatics and Telecommunications, National and Kapodistrian University of Athens Campus, Ilisia, 15784 Athens, Greece; kostas@chatzi.org

[2] Department of Computing, Macquarie University, Ryde City 2109, Australia; tashfernandes@gmail.com

[3] Inria and Institut Polytechnique de Paris, 91120 Palaiseau, France

* Correspondence: catuscia@lix.polytechnique.fr

† This paper is an extended version of our paper by Chatzikokolakis, K.; Fernandes, N.; Palamidessi, C. Comparing systems: max-case refinement orders and application to differential privacy. In Proceedings of the 32nd IEEE Computer Security Foundations Symposium, Hoboken, NJ, USA, 25–28 June 2019.

Received: 27 October 2020; Accepted: 7 December 2020; Published: 12 December 2020

Abstract: Quantitative Information Flow (QIF) and Differential Privacy (DP) are both concerned with the protection of sensitive information, but they are rather different approaches. In particular, QIF considers the expected probability of a successful attack, while DP (in both its standard and local versions) is a max-case measure, in the sense that it is compromised by the existence of a possible attack, regardless of its probability. Comparing systems is a fundamental task in these areas: one wishes to guarantee that replacing a system A by a system B is a safe operation that is the privacy of B is no worse than that of A. In QIF, a refinement order provides strong such guarantees, while, in DP, mechanisms are typically compared w.r.t. the privacy parameter ε in their definition. In this paper, we explore a variety of refinement orders, inspired by the one of QIF, providing precise guarantees for max-case leakage. We study simple structural ways of characterising them, the relation between them, efficient methods for verifying them and their lattice properties. Moreover, we apply these orders in the task of comparing DP mechanisms, raising the question of whether the order based on ε provides strong privacy guarantees. We show that, while it is often the case for mechanisms of the same "family" (geometric, randomised response, etc.), it rarely holds across different families.

Keywords: quantitative information flow; differential privacy; security refinement orderings

1. Introduction

The enormous growth in the use of internet-connected devices and the big-data revolution have created serious privacy concerns, and motivated an intensive area of research aimed at devising methods to protect the users' sensitive information. During the last decade, two main frameworks have emerged in this area: Differential Privacy (DP) and Quantitative Information Flow (QIF).

Differential privacy (DP) [1] was originally developed in the area of statistical databases, and it aims at protecting the individuals' data while allowing the release of aggregate information through queries. This is obtained by *obfuscating* the result of the query via the addition of controlled noise. Naturally, we need to assume that the *curator*, namely the entity collecting and storing the data and handling the queries, is honest and capable of protecting the data from security breaches. Since this assumption cannot always be guaranteed, a variant has been proposed: local differential privacy (LDP) [2], where the data are obfuscated individually before they are collected.

Both DP and LPD are subsumed by d-privacy [3], and, in this paper, we will use the latter as a unifying framework. The definition of d-privacy assumes an underlying metric structure on the data domain $\mathcal{X}$. An obfuscation mechanism K for $\mathcal{X}$ is a probabilistic mapping from $\mathcal{X}$ to some output domain $\mathcal{Y}$, namely a function from $\mathcal{X}$ to probabilistic distributions over $\mathcal{Y}$. We will use the notation $K_{x,y}$ to represent the probability that K on input x gives output y. The mechanism K is $\varepsilon{\cdot}d$-private, where ε is a parameter representing the privacy level, if

$$K_{x_1,y} \leq e^{\varepsilon\, d(x_1,x_2)} K_{x_2,y} \qquad \text{for all } x_1, x_2 \in \mathcal{X}, y \in \mathcal{Y} \tag{1}$$

Standard DP is obtained from this definition by assuming $\mathcal{X}$ to be a set of all datasets and d the Hamming distance between two datasets, seen as vectors or records (i.e., the number of positions in which the two datasets differ). Note that the more common definition of differential privacy assumes that x_1, x_2 are adjacent, i.e., their Hamming distance is 1, and requires $K_{x_1,y} \leq e^{\varepsilon}\ K_{x_2,y}$. It is easy to prove that the two definitions are equivalent. As for LDP, it is obtained by considering the so-called *discrete metric* which assigns distance 0 to identical elements, and 1 otherwise.

The other framework for the protection of sensitive information, quantitative information flow (QIF), focuses on the potentialities and the goals of the attacker, and the research on this area has developed rigorous foundations based on information theory [4,5]. The idea is that a system processing some sensitive data from a random variable X and releasing some observable data as a random variable Y can be modelled as an information-theoretic channel with input X and output Y. The leakage is then measured in terms of correlation between X and Y. There are, however, many different ways to define such correlation, depending on the notion of adversary. In order to provide a unifying approach, Ref. [6] has proposed the theory of g-leakage, in which an adversary is characterized by a functional parameter g representing its *gain* for each possible outcomes of the attack.

One issue that arises in both frameworks is how to compare systems from the point of view of their privacy guarantees. It is important to have a rigorous and effective way to establish whether a mechanism is better or worse than another one, in order to guide the design and the implementation of mechanisms for information protection. This is not always an obvious task. To illustrate the point, consider the following examples.

Example 1. *Let P_1, P_2, P_3 and P_4 be the programs illustrated in Table 1, where* H *is a "high" (i.e., secret) input and* L *is a "low" (i.e., public) output. We assume that* H *is a uniformly distributed* 32*-bit integer with range* $0 \leq \mathtt{H} < 2^{32}$. *All these programs leak information about* H *via* L*, in different ways: P_1 reveals* H *whenever it is a multiple of* 8 *(*H mod 8 *represents the integer division of* H *by* 8*), and reveals nothing otherwise. P_2 does the same thing whenever* H *is a multiple of* 4*. P_3 reveals the last* 8 *bits of* H *(note that* $\mathtt{H}\ \&\ 0^{24}1^{8}$ *represents the bitwise conjunction between* H *and a string of* 24 *bits "*0*" followed by* 8 *bits "*1*"). Analogously, P_4 reveals the last* 4 *bits of* H*. Now, it is clear that P_2 leaks more than P_1, and that P_4 leaks more than P_3, but how to compare, for instance, P_1, and P_3? It is debatable which one is worse because their behavior is very different: P_1 reveals nothing in most cases, but when it does reveal something, it reveals everything. P_3, on the other hand, always reveals part of the secret. Clearly, we cannot decide which situation is worse, unless we have some more information about the goals and the capabilities of the attacker. For instance, if the adversary has only one attempt at his disposal (and no extra information), then the program P_3 is better because even after the output of* L *there are still* 24 *bits of* H *that are unknown. On the other hand, if the adversary can repeat the attacks on program similar to P_3, then eventually it will uncover the secret entirely all the times.*

Table 1. Programs that take in input a secret H and leak information about H via the output L.

P_1	P_2	P_3	P_4

`if H mod 8 = 0 then`
`L := H`
`else`
`L := 1`
	`if H mod 4 = 0 then`
	`L := H`
	`else`
	`L := 1`
		`L := H &` $0^{24}1^{8}$	`L := H &` $0^{28}1^{4}$

Example 2. *Consider the domain of the integer numbers between* 0 *and* 100*, and consider a geometric mechanism (cf. Definition 9) on this domain, with* $\varepsilon = \log {}^{27}\!/_{25}$*. Then, consider a randomized response mechanism (cf. Definition 13) still on the same domain and* $\varepsilon = \log 2$*. The two mechanisms are illustrated in Figure 1. They both satisfy d-privacy, but for different d: in the first case d is the standard distance between numbers, while in the second case it is the discrete metric. Clearly, it does not make sense to compare these mechanisms on the basis of their respective privacy parameters ε because they represent different privacy properties, and it is not obvious how to compare them in general: The geometric mechanism tends to make the true value indistinguishable from his immediate neighbors, but it separates it from the values far away. The randomized response introduces the same level of confusion between the true value and any other value of the domain. Thus, which mechanism is more private depends on the kind of attack we want to mitigate: if the attacker is trying to guess an approximation of the value, then the randomized response is better. If the attacker is only interested in identifying the true value among the immediate neighbors, then the geometric is better. Indeed, note that in the subdomain of the numbers between* 40 *and* 60 *the geometric mechanism also satisfies d-privacy with d being the discrete metric and* $\varepsilon = \log 2$*, and in any subdomain smaller than that it satisfies the discrete metric d-privacy with* $\varepsilon < \log 2$*.*

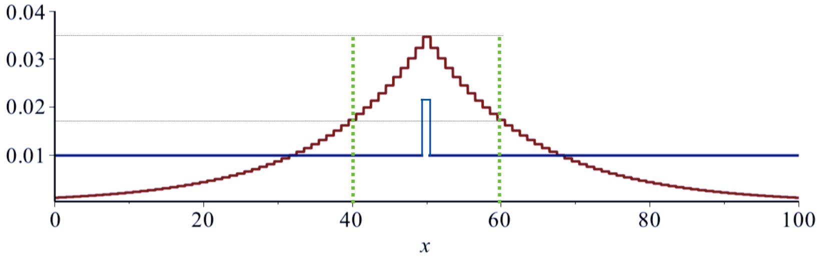


Figure 1. Comparison between the geometric (**red**) and the randomized response (**blue**) mechanisms. The area between 40 and 60, delimited by the green lines, represents the sub-domain where the geometric mechanism satisfies also the discrete metric *d*-privacy with $\varepsilon = \log 2$.

In this respect, the QIF approach has led to an elegant theory of refinement (pre)order (In this paper, we call $\sqsubseteq_{\mathbb{G}}^{\text{avg}}$ and the other refinement relations "orders", although, strictly speaking, they are preorders.) $\sqsubseteq_{\mathbb{G}}^{\text{avg}}$, which provides strong guarantees: $A \sqsubseteq_{\mathbb{G}}^{\text{avg}} B$ means that B *is safer than* A *in all circumstances*, in the sense that the *expected gain* of an attack on B is less than on A, for whatever kind of gain the attacker may be seeking. This means that we can always substitute the component A by B without compromising the security of the system. An appealing aspect of this particular refinement order is that it is characterized by a precise structural relation between the stochastic channels associated with A and B [6,7], which makes it easy to reason about, and relatively efficient to verify. It is important to remark that this order is based on an *average* notion of adversarial gain (*vulnerability*), defined by mediating over all possible observations and their probabilities. We call this perspective *average-case*.

At the other end of the spectrum, DP, LDP and *d*-privacy are *max-case* measures. In fact, by applying the Bayes theorem to (1), we obtain:

$$\frac{p(x_1 \mid y)}{p(x_2 \mid y)} \le e^{\varepsilon\, d(x_1,x_2)} \frac{\pi(x_1)}{\pi(x_2)} \qquad \text{for all } x_1, x_2 \in \mathcal{X}, y \in \mathcal{Y} \tag{2}$$

where, for $i \in \{1,2\}$, $\pi(x_i)$ is the *prior* probability of x_i and $p(x_i \mid y)$ is the *posterior* probability of x_i given y. We can interpret ${}^{\pi(x_1)}\!/_{\pi(x_2)}$ and ${}^{p(x_1|y)}\!/_{p(x_2|y)}$ as knowledge about $\mathcal{X}$: they represent how much more likely x_1 is with respect to x_2, *before* (prior) and *after* (posterior) observing y, respectively. Thus, the property expresses a bound on how much the adversary can learn from each individual outcome of the mechanism (The property (2) is also the semantic interpretation of the guarantees of the Pufferfish framework (cf. [8], Section 3.1). The ratio $\frac{p(x_1|y)}{p(x_2|y)} / \frac{\pi(x_1)}{\pi(x_2)}$ is known as *odds ratio*).

In the literature of DP, LDP and d-privacy, mechanisms are usually compared on the basis of their ε-value (In DP and LDP ε is a parameter that usually appears explicitly in the definition of the mechanism. In d-privacy, it is an implicit scaling factor.), which controls a bound on the log-likelihood ratio of an observation y given two "secrets" x_1 and x_2: smaller ε means more privacy. In DP and LDP, the bound is ε itself, while in d-privacy it is $\varepsilon \times d(s_1, s_2)$ We remark that the relation induced by ε in d-privacy is fragile, in the sense that the definition of d-privacy assumes an underlying metric structure d on the data, and whether a mechanism B is "better" than A depends in general on the metric considered.

Average-case and max-case are different principles, suitable for different scenarios: the former represents the point of view of an organization, for instance an insurance company providing coverage for risks related to credit cards, which for the cost–benefit analysis is interested in reasoning in terms of expectation (expected cost of an attack). The max-case represents the point of view of an individual, who is interested in limiting the cost of *any* attack. As such, the max-case seems particularly suitable for the domain of privacy.

In this paper, we combine the max-case perspective with the robustness of the QIF approach, and we introduce two refinement orders:

- $\sqsubseteq^{\max}_{\mathbb{Q}}$, based on the max-case leakage introduced in [9]. This order takes into account all possible privacy breaches caused by any observable (like in the DP world), but it quantifies over all possible quasi-convex vulnerability functions (in the style of the QIF world).
- $\sqsubseteq^{\mathrm{prv}}_{\mathbb{M}}$, based on d-privacy (like in the DP world), but quantified over all metrics d.

To underline the importance of a robust order, let us consider the case of the oblivious mechanisms for differential privacy: These mechanisms are of the form $K = H \circ f$, where $f : \mathcal{X} \to \mathcal{Y}$ is a query, namely a function from datasets in $\mathcal{X}$ to some answer domain $\mathcal{Y}$, and H is a probabilistic mechanism implementing the noise. The idea is that the system first computes the result $y \in \mathcal{Y}$ of the query (*true answer*), and then it applies H to y to obtain a *reported answer* z. In general, if we want K to be ε-DP, we need to tune the mechanism H in order to take into account the *sensitivity* of f, which is the maximum distance between the results of f on two adjacent databases, and as such it depends on the metric on $\mathcal{Y}$. However, if we know that $K = H \circ f$ is ε-DP, and that $H \sqsubseteq^{\mathrm{prv}}_{\mathbb{M}} H'$ for some other mechanism H', then we can safely substitute H by H' as it is because one of our results (cf. Theorem A5) guarantees that $K' = H' \circ f$ is also ε-DP. In other words, $H \sqsubseteq^{\mathrm{prv}}_{\mathbb{M}} H'$ implies that we can substitute H by H' in an oblivious mechanism for whatever query f and whatever metric on $\mathcal{Y}$, without the need to know the sensitivity of f and without the need to do any tuning of H' . Thanks to Theorems 3 and 4, we know that this is the case also for $\sqsubseteq^{\mathrm{avg}}_{\mathbb{G}}$ and $\sqsubseteq^{\max}_{\mathbb{Q}}$. We illustrate this with the following example.

Example 3. *Consider datasets $x \in \mathcal{X}$ of records containing the age of people, expressed as natural numbers from* 0 *to* 100*, and assume that each dataset in $\mathcal{X}$ contains at least* 100 *records. Consider two queries, $f(x)$ and $g(x)$, which give the rounded average age and the minimum age of the people in x, respectively. Finally, consider the* truncated geometric *mechanism TG^{ε} (cf. Definition 10), and the randomized response mechanism R^{ε} (cf. Definition 13). It is easy to see that $K_1 = TG^{\varepsilon} \circ f$ is ε-DP, and it is possible to prove that $TG^{\varepsilon} \sqsubseteq^{\mathrm{prv}}_{\mathbb{M}} R^{\varepsilon}$ (cf. Theorem 14). We can then conclude that $K_2 = R^{\varepsilon} \circ f$ is ε-DP as well, and that in general it is safe to replace TG^{ε} by R^{ε} for whatever query. On the other hand, $R^{\varepsilon} \not\sqsubseteq^{\mathrm{prv}}_{\mathbb{M}} TG^{\varepsilon}$, so we cannot expect that it is safe to replace R^{ε} by TG^{ε} in any context. In fact, $K_3 = R^{\varepsilon} \circ g$ is ε-DP, but $K_4 = TG^{\varepsilon} \circ g$* is not *$\varepsilon$-DP, despite the fact that both mechanisms are constructed using the same privacy parameter ε. Hence, we can conclude that a refinement relation based only on the comparison of the ε parameters would not be robust, at least not for a direct replacement in an arbitrary context. Note that K_4 is* 100 $\times$ *ε-DP. In order to make it ε-DP, we should divide the parameter ε by the sensitivity of g (with respect to the ordinary distance on natural numbers), which is* 100*, i.e., use $TG^{\varepsilon/100}$. For R^{ε}, this is not necessary because it is defined using the discrete metric on $\{0, \ldots, 100\}$, and the sensitivity of g with respect to this metric is* 1.

The robust orders allow us to take into account different kinds of adversaries. The following example shows what the idea is.

Example 4. *Consider the following three LDP mechanisms, represented by their stochastic matrices (where each element is the conditional probability of the outcome of the mechanism, given the secret value). The secrets are three possible economic situations of an individual, p, a and r, standing for "poor", "average" and "rich", respectively. The observable outcomes are r and n, standing for "rich" and "not rich".*

A	n	r
p	$3/4$	$1/4$
a	$1/2$	$1/2$
r	$1/4$	$3/4$

B	n	r
p	$2/3$	$1/3$
a	$2/3$	$1/3$
r	$1/3$	$2/3$

C	n	r
p	$2/3$	$1/3$
a	$1/2$	$1/2$
r	$1/3$	$2/3$

(3)

Let us assume that the prior distribution π on the secrets is uniform. We note that A is $(\log 3)$*-LDP while B is* $(\log 2)$*-LDP. Hence, if we only look at the value of ε, we would think that B is better than A from the privacy point of view. However, there are attackers that gain more from B than from A (which means that, with respect to those attackers, the privacy of B is worse). For instance, this is the case when the attacker is only interested in discovering whether the person is* rich *or not. In fact, if we consider a gain* 1 *when the attacker guesses the right class (r versus (either p or a)) and* 0 *otherwise, we have that the highest possible gain in A is* $(3/4)\,\pi(p) + (1/2)\,\pi(a) = 5/12$*, while in B is* $(2/3)\,\pi(p) + (2/3)\,\pi(a) = 4/9$*, which is higher than* $5/12$*. This is consistent with our orders: it is possible to show that none of the three orders hold between A and B, and that therefore we should not expect B to be better (for privacy) than A with respect to all possible adversaries.*

On the other hand, the mechanism C is also $(\log 2)$*-LDP, and in this case we have that the relation* $A \sqsubseteq_{\mathbb{Q}}^{\max} C$ *holds, implying that we can safely replace A by C. We can also prove that the reverse does not hold, which means that C is strictly better than A.*

A fundamental issue is how to prove that these robust orders hold: Since $\sqsubseteq_{\mathbb{Q}}^{\max}$ and $\sqsubseteq_{\mathbb{M}}^{\mathrm{prv}}$ involve universal quantifications, it is important to devise finitary methods to verify them. To this purpose, we will study their characterizations as structural relations between stochastic matrices (representing the mechanisms to be compared), along the lines of what was done for $\sqsubseteq_{\mathbb{G}}^{\mathrm{avg}}$.

We will also study the relation between the three orders (the two above and $\sqsubseteq_{\mathbb{G}}^{\mathrm{avg}}$), and their algebraic properties. Finally, we will analyze various mechanisms for DP, LDP, and d-privacy to see in which cases the order induced by ε is consistent with the three orders above.

1.1. Contribution

The main contributions of this paper are the following:

- We introduce two refinement orders for the max case, $\sqsubseteq_{\mathbb{Q}}^{\max}$ and $\sqsubseteq_{\mathbb{M}}^{\mathrm{prv}}$, that are robust with respect to a large class of adversaries.
- We give structural characterizations of both $\sqsubseteq_{\mathbb{Q}}^{\max}$ and $\sqsubseteq_{\mathbb{M}}^{\mathrm{prv}}$ in terms of relations on the stochastic matrices of the mechanisms under comparison. These relations help the intuition and open the way to verification.
- We study efficient methods to verify the structural relations above. Furthermore, these methods are such that, when the verification fails, they produce counterexamples. In this way, it is possible to pin down what the problem is and try to correct it.
- We show that $\sqsubseteq^{\mathrm{avg}} \subset \sqsubseteq^{\max} \subset \sqsubseteq^{\mathrm{prv}}$.
- We apply the three orders ($\sqsubseteq_{\mathbb{G}}^{\mathrm{avg}}$, $\sqsubseteq_{\mathbb{Q}}^{\max}$, and $\sqsubseteq_{\mathbb{M}}^{\mathrm{prv}}$) to the comparison of some well-known families of d-private mechanisms: geometric, exponential and randomised response. We show that, in general, $A \sqsubseteq_{\mathbb{G}}^{\mathrm{avg}} B$ (and thus all the refinement orders between A and B) holds within the same family whenever the ε of B is smaller than that of A.

- We show that, if A and B are mechanisms from different families, then, even if the ε of B is smaller than that of A, the relations $A \sqsubseteq_{\mathbb{G}}^{\text{avg}} B$ and $A \sqsubseteq_{\mathbb{Q}}^{\max} B$ do not hold, and in most cases $A \sqsubseteq_{\mathbb{M}}^{\text{prv}} B$ does not hold either. We conclude that a comparison based only on the value of the ε's is not robust across different families, at least not for the purposes illustrated above.
- We study lattice-properties of these orders. In contrast to $\sqsubseteq_{\mathbb{G}}^{\text{avg}}$, which was shown to not be a lattice, we prove that suprema and infima exist for $\sqsubseteq_{\mathbb{Q}}^{\max}$ and $\sqsubseteq_{\mathbb{M}}^{\text{prv}}$, and that therefore these orders form lattices.

1.2. Related Work

We are not aware of many studies on refinement relations for QIF. Yasuoka and Terauchi [10] and Malacaria [11] have explored strong orders on *deterministic* mechanisms, focusing on the fact that such mechanisms induce *partitions* on the space of secrets. They showed that the orders produced by min-entropy leakage [5] and Shannon leakage [12,13] are the same and, moreover, they coincide with the *partition refinement* order in the *Lattice of Information* [14]. This order was extended to the probabilistic case in [6], resulting in the relation $\sqsubseteq^{\text{avg}}$ mentioned in Section 2. The same paper [6] proposed the theory of g-leakage and introduced the corresponding order $\sqsubseteq_{\mathbb{G}}^{\text{avg}}$. Furthermore, [6] proved that $\sqsubseteq^{\text{avg}} \subseteq \sqsubseteq_{\mathbb{G}}^{\text{avg}}$ and conjectured that also the reverse should hold. This conjecture was then proved valid in [7]. The max-case leakage, on which the relation $\sqsubseteq_{\mathbb{Q}}^{\max}$ is based, was introduced in [9], but $\sqsubseteq_{\mathbb{Q}}^{\max}$ and its properties were not investigated. Finally, $\sqsubseteq^{\text{prv}}$ is a novel notion introduced in this paper.

In the field of differential privacy, on the other hand, there have been various works aimed at trying understand the operational meaning of the privacy parameter ε and at providing guidelines for the choice of its values. We mention, for example [15,16], which consider the value of ε from an economical point of view, in terms of cost. We are not aware, however, of studies aimed at establishing orders between the level of privacy of different mechanisms, except the one based on the comparison of the ε's.

The relation between QIF and DP, LDP, and d-privacy is based on the so-called *semantic interpretation* of the privacy notions that regard these properties as expressing a bounds on the increase of knowledge (from prior to posterior) due to the answer reported by the mechanism. For d-privacy, the semantic interpretation is expressed by (2). To the best of our knowledge, this interpretation was first pointed out (for the location privacy instance) in [17]. The seminal paper on d-privacy, [3], also proposed a semantic interpretation, with a rather different flavor, although formally equivalent. As for DP, as explained in the Introduction, (2) instantiated to databases and Hamming distance corresponds to the odds ratio on which the semantics interpretation is based provided in [8]. Before that, another version of semantic interpretation was presented in [1] and proved equivalent to a form of DP called ε-indistinguishability. Essentially, in this version, an adversary that queries the database, and knows all the database except one record, cannot infer too much about this record from the answer to the query reported by the mechanism. Later on, an analogous version of semantic interpretation was reformulated in [18] and proved equivalent to DP. A different interpretation of DP, called *semantic privacy*, was proposed by [19]. This interpretation is based on a comparison between two posteriors (rather between the posterior and the prior), and the authors show that, within certain limits, it is equivalent to DP.

A short version of this paper, containing only some of the proofs, appeared in [20].

1.3. Plan of the Paper

In the next three sections, Sections 2–4, we define the order refinements $\sqsubseteq_{\mathbb{G}}^{\text{avg}}$, $\sqsubseteq_{\mathbb{Q}}^{\max}$ and $\sqsubseteq_{\mathbb{M}}^{\text{prv}}$, respectively, and we study their properties. In Section 5.1, we investigate methods to verify them. In Section 6, we consider various mechanisms for DP and its variants, and we investigate the relation between the parameter ε and the orders introduced in this paper. In Section 7, we show that $\sqsubseteq_{\mathbb{Q}}^{\max}$ and $\sqsubseteq_{\mathbb{M}}^{\text{prv}}$ form a lattice. Finally, Section 8 provides conclusions.

Note: In order to make the reading of the paper more fluid, we have moved all the proofs to the appendix at the end of the paper.

2. Average-Case Refinement

We recall here some basic concepts from the literature. Table 2 lists the symbols used for the main concepts through the paper.

Table 2. Definitions and symbols used in this paper.

Defn	Vulnerabilities
Equation (4)	$V[\pi, C] := \sum_y a_y V(\delta^y)$
Section 3	$V^{\max}[\pi, C] := \max_y V(\delta^y)$
Definition 8	$V_d(\pi) := \inf\{\varepsilon \geq 0 \mid \forall x, x' \in \mathcal{X},\ \pi_x \leq e^{\varepsilon \cdot d(x,x')} \pi_{x'}\}$
Defn	**Leakage Measures**
Definition 4	$\mathrm{Priv}_d(C) := \inf\{\varepsilon \geq 0 \mid C \text{ satisfies } \varepsilon \cdot d\text{-privacy}\}$
Equation (23)	$\mathcal{L}_d^{+,\max}(\pi, C) := V_d^{\max}[\pi, C] - V_d(\pi)$
Equation (24)	$\mathcal{ML}_d^{+,\max}(C) := \max_\pi \mathcal{L}_d^{+,\max}(\pi, C)$
Defn	**Refinement Orders**
Equation (6)	$A \sqsubseteq^{\mathrm{avg}} B$ iff $AR = B$ for some channel R
Definition 2	$A \sqsubseteq^{\max} B$ iff $R\tilde{A} = \tilde{B}$ for some channel R
Definition 7	$A \sqsubseteq^{\mathrm{prv}} B$ iff B satisfies d_A-privacy
Defn	**Leakage Orders**
Section 2.2	$A \sqsubseteq_{\mathbb{G}}^{\mathrm{avg}} B$ iff $\forall g\colon \mathbb{G}\mathcal{X}, \forall \pi\colon \mathbb{D}\mathcal{X},\ V_g[\pi, A] \geq V_g[\pi, B]$
Definition 1	$A \sqsubseteq_{\mathbb{Q}}^{\max} B$ iff $\forall V\colon \mathbb{Q}\mathcal{X}, \forall \pi\colon \mathbb{D}\mathcal{X},\ V^{\max}[\pi, A] \ \geq\ V^{\max}[\pi, B]$
Definition 6	$A \sqsubseteq_{\mathbb{M}}^{\mathrm{prv}} B$ iff $\forall d \in \mathbb{M}\mathcal{X},\ A \sqsubseteq_d^{\mathrm{prv}} B$
Definition 5	$A \sqsubseteq_d^{\mathrm{prv}} B$ iff $\mathrm{Priv}_d(A) \geq \mathrm{Priv}_d(B)$

2.1. Vulnerability, Channels, and Leakage

Quantitative Information Flow studies the problem of quantifying the *information leakage* of a system (e.g., a program, or an anonymity protocol). A common model in this area is to consider that the user has a secret x from a finite set of possible secrets $\mathcal{X}$, about which the adversary has some probabilistic knowledge $\pi\colon \mathbb{D}\mathcal{X}$ ($\mathbb{D}\mathcal{X}$ denoting the set of probability distributions over $\mathcal{X}$). A function $V\colon \mathbb{D}\mathcal{X} \to \mathbb{R}_{\geq 0}$ is then employed to measure the *vulnerability* of our system: $V(\pi)$ quantifies the adversary's success in achieving some desired goal, when his knowledge about the secret is π.

Various such functions can be defined (e.g., employing well-known notions of entropy), but it quickly becomes apparent that no single vulnerability function is meaningful for all systems. The family of g-vulnerabilities [6] tries to address this issue by parametrizing V in an operational scenario: first, the adversary is assumed to possess a set of *actions* $\mathcal{W}$; second, a gain function $g(w, x)$ models the adversary's gain when choosing action w and the real secret is x. g-vulnerability can be then defined as the expected gain of an optimal guess: $V_g(\pi) = \max_{w\colon \mathcal{W}} \sum_{x\colon \mathcal{X}} \pi_x g(w, x)$. Different adversaries can be modelled by proper choices of $\mathcal{W}$ and g. We denote by $\mathbb{G}\mathcal{X}$ the set of all gain functions.

A system is then modelled as a *channel*: a probabilistic mapping from the (finite) set of secrets $\mathcal{X}$ to a finite set of observations $\mathcal{Y}$, described by a stochastic matrix C, where $C_{x,y}$ is the probability that secret x produces the observation y. When the adversary observes y, he can transform his initial knowledge π into a *posterior* knowledge $\delta^y\colon \mathbb{D}\mathcal{X}$. Since each observation y is produced with some probability a_y, it is sometimes conceptually useful to consider that the *result of running a channel* C, on the initial knowledge π, is a "hyper" distribution $[\pi, C]$: a probability distribution on posteriors δ^y, each having probability a_y.

It is then natural to define the (average-case) *posterior vulnerability* of the system by applying V to each posterior δ^y, then averaging by its probability a_y of being produced:

$$V[\pi, C] \quad := \quad \textstyle\sum_y a_y V(\delta^y) \tag{4}$$

when defining vulnerability in this way, it can be shown [9] that V has to be *convex on* π; otherwise, fundamental properties (such as the data processing inequality) are violated. Any continuous and convex function V can be written as V_g for a properly chosen g, so, when studying average-case leakage, we can safely restrict to using g-vulnerability.

Leakage can be finally defined by comparing the prior and posterior vulnerabilities, e.g., as $\mathcal{L}_g^+(\pi, C) = V_g[\pi, C] - V_g(\pi)$. (Comparing vulnerabilities "multiplicatively" is also possible but is orthogonal to the goals of this paper.)

2.2. Refinement

A fundamental question arises in the study of leakage: can we guarantee that a system B is no less safe than a system A? Having a family of vulnerability functions, we can naturally define a strong order $\sqsubseteq_{\mathbb{G}}^{\text{avg}}$ on channels by explicitly requiring that B leaks (Note that comparing the leakage of A, B is equivalent to comparing their posterior vulnerability, so we choose the latter for simplicity.) no more than A, for all priors π and all gain functions $g\colon \mathbb{G}\mathcal{X}$: (Note also that quantifying over $g\colon \mathbb{G}\mathcal{X}$ is equivalent to quantifying over all continuous and convex vulnerabilities.)

$$A \sqsubseteq_{\mathbb{G}}^{\text{avg}} B \qquad \text{iff} \qquad V_g[\pi, A] \geq V_g[\pi, B] \qquad\qquad \text{for all } g\colon \mathbb{G}\mathcal{X}, \pi\colon \mathbb{D}\mathcal{X} \tag{5}$$

Although $\sqsubseteq_{\mathbb{G}}^{\text{avg}}$ is intuitive and provides clear leakage guarantees, the explicit quantification over vulnerability functions makes it hard to reason about and verify. Thankfully, this order can be characterized in a "structural" way that is as a direct property of the channel matrix. We first define the *refinement* order $\sqsubseteq^{\text{avg}}$ on channels by requiring that B can be obtained by post-processing A by some other channel R that is:

$$A \sqsubseteq^{\text{avg}} B \qquad \text{iff} \qquad AR = B \ \text{ for some channel R} \tag{6}$$

A fundamental result [6,7] states that $\sqsubseteq^{\text{avg}}$ and $\sqsubseteq_{\mathbb{G}}^{\text{avg}}$ coincide.

We read $A \sqsubseteq^{\text{avg}} B$ as "A is refined by B", or "B is as safe as A". When $A \sqsubseteq^{\text{avg}} B$ holds, we have a strong privacy guarantee: we can safely replace A by B without decreasing the privacy of the system, independently from the adversary's goals and his knowledge. However, refinement can be also useful in case $A \not\sqsubseteq^{\text{avg}} B$; namely, we can conclude that some adversary must exist, modelled by a gain function g, and some initial knowledge π, such that the adversary actually prefers to interact with A rather than interacting with B. (Whether this adversary is of practical interest or not is a different issue, but we know that one exists.) Moreover, we can actually *construct* such a "counter-example" gain function; this is discussed in Section 5.

3. Max-Case Refinement

Although $\sqsubseteq^{\text{avg}}, \sqsubseteq_{\mathbb{G}}^{\text{avg}}$ provides a strong and precise way of comparing systems, one could argue that average-case vulnerability might underestimate the threat of a system. More precisely, imagine that there is a certain observation y such that the corresponding posterior δ^y is highly vulnerable (e.g., the adversary can completely infer the real secret), but y happens with very small probability a_y.

In this case, the average-case posterior vulnerability $V[\pi, C]$ can be relatively small, although $V(\delta^y)$ is large for that particular y.

If such a scenario is considered problematic, we can naturally quantify leakage using a max-case (called "worse"-case in some contexts, although the latter is more ambiguous, "worse" can refer to a variety of factors.) variant of posterior vulnerability, where all observations are treated equally regardless of their probability of being produced:

$$V^{\max}[\pi, C] \quad := \quad \max_y V(\delta^y) \tag{7}$$

Under this definition, it can be shown [9] that V has to be *quasi-convex* on π (instead of convex), in order to satisfy fundamental properties (such as the data processing inequality). Hence, in the max-case, we no longer restrict to g-vulnerabilities (which are always convex), but we can use any vulnerability $V : \mathbb{Q}\mathcal{X}$, where $\mathbb{Q}\mathcal{X}$ denotes the set of all continuous quasi-convex functions $\mathbb{D}\mathcal{X} \to \mathbb{R}_{\geq 0}$.

Inspired by $\sqsubseteq_{\mathbb{G}}^{\text{avg}}$, we can now define a corresponding max-case leakage order.

Definition 1. *The max-case leakage order is defined as*

$$A \sqsubseteq_{\mathbb{Q}}^{\max} B \qquad \textit{iff} \qquad V^{\max}[\pi, A] \ \geq \ V^{\max}[\pi, B] \qquad\qquad \text{for all } V{:}\,\mathbb{Q}\mathcal{X}, \pi{:}\,\mathbb{D}\mathcal{X} \tag{8}$$

Similarly to its average-case variant, $\sqsubseteq_{\mathbb{Q}}^{\max}$ provides clear privacy guarantees by explicitly requiring that B leaks no more than A for all adversaries (modelled as a vulnerability V). However, this explicit quantification makes the order hard to reason about and verify. We would thus like to characterize $\sqsubseteq_{\mathbb{Q}}^{\max}$ by a refinement order that depends only on the structure of the two channels.

Given a channel C from $\mathcal{X}$ to $\mathcal{Y}$, we denote by $\tilde{C}$ the channel obtained by normalizing C's columns (if a column consists of only zeroes, it is simply removed.) and then transposing:

$$\tilde{C}_{y,x} \quad := \quad \frac{C_{x,y}}{\sum_x C_{x,y}} \tag{9}$$

Note that the row y of $\tilde{C}$ can be seen as the posterior distribution δ^y obtained by C under the uniform prior. Note also that $\tilde{C}$ is non-negative and its rows sum up to 1, so it is a valid channel from $\mathcal{Y}$ to $\mathcal{X}$. The average-case refinement order required that B can be obtained by post-processing A. We define the *max-case refinement* order by requiring that $\tilde{B}$ can be obtained by *pre-processing* $\tilde{A}$.

Definition 2. *The max-case refinement order is defined as $A \sqsubseteq^{\max} B$ iff $R\tilde{A} = \tilde{B}$ for some channel R.*

Our goal now is to show that $\sqsubseteq_{\mathbb{Q}}^{\max}$ and $\sqsubseteq^{\max}$ are different characterizations of the same order. To do so, we start by giving a "semantic" characterization of $\sqsubseteq^{\max}$ that is, expressing it, not in terms of the channel matrices A and B, but in terms of the *posterior distributions* that they produce. Thinking of $[\pi, C]$ as a ("hyper") distribution on the posteriors produced by π and C, its support *supp* $[\pi, C]$ is the set of all posteriors produced with non-zero probability. We also denote by *ch* S the convex hull of S.

Theorem 1. *Let $\pi{:}\,\mathbb{D}\mathcal{X}$. If $A \sqsubseteq^{\max} B$, then the posteriors of B (under π) are convex-combinations of those of A, that is,*

$$\mathit{supp}\,[\pi, B] \quad \subseteq \quad \mathit{ch}\,\mathit{supp}\,[\pi, A] \tag{10}$$

Moreover, if (A1) *holds and π is full support, then $A \sqsubseteq^{\max} B$.*

Note that, if (A1) holds for any full-support prior, then it must hold for all priors.

Theorem A1 has a nice geometric intuition (cf. Figure 2) that we are going to illustrate in the following example.

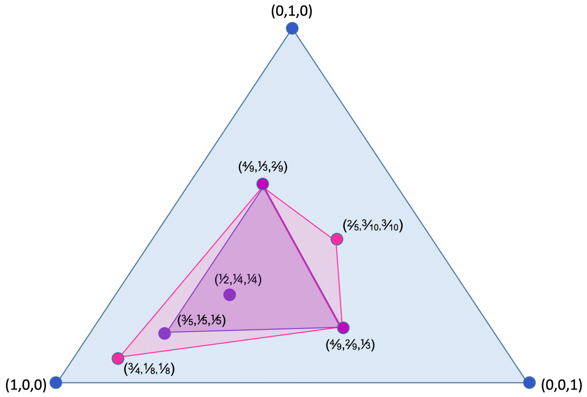


Figure 2. The simplex and the convex hulls of the posterior distributions in Example 5.

Example 5. *Consider the following systems.*

$$
\begin{array}{|c|cccc|}
\hline
A & y_1 & y_2 & y_3 & y_4 \\
\hline
x_1 & 1/3 & 2/9 & 2/9 & 2/9 \\
x_2 & 1/9 & 1/3 & 2/9 & 1/3 \\
x_3 & 1/9 & 2/9 & 1/3 & 1/3 \\
\hline
\end{array}
\qquad
\begin{array}{|c|cccc|}
\hline
B & y_1 & y_2 & y_3 & y_4 \\
\hline
x_1 & 1/3 & 2/9 & 2/9 & 1/9 \\
x_2 & 2/9 & 1/3 & 2/9 & 1/9 \\
x_3 & 2/9 & 2/9 & 1/3 & 1/9 \\
\hline
\end{array}
\tag{11}
$$

Consider the prior $\pi = (1/2, 1/4, 1/4)$. The set of the posterior distributions generated by A under π are:

$$
supp\,[\pi, A] = \{(3/4, 1/8, 1/8), (4/9, 1/3, 2/9), (4/9, 2/9, 1/3), (2/5, 3/10, 3/10)\} \tag{12}
$$

while those generated by B are:

$$
supp\,[\pi, B] = \{(3/5, 1/5, 1/5), (4/9, 1/3, 2/9), (4/9, 2/9, 1/3), (1/2, 1/4, 1/4)\} \tag{13}
$$

These posteriors, and the convex hulls that they generate, are illustrated in Figure 2. The pink area is the $ch\,supp\,[\pi, A]$ and the purple area is the $ch\,supp\,[\pi, B]$. We can see that $supp\,[\pi, B] \subseteq ch\,supp\,[\pi, A]$, or, equivalently, $ch\,supp\,[\pi, B] \subseteq ch\,supp\,[\pi, A]$.

We are now ready to give the main result of this section.

Theorem 2. *The orders $\sqsubseteq^{\max}$ and $\sqsubseteq^{\max}_{\mathbb{Q}}$ coincide.*

Similarly to the average case, $A \sqsubseteq^{\max} B$ gives us a strong leakage guarantee: can safely replace A by B, knowing that, for any adversary, the max-case leakage of B can be no-larger than that of A. Moreover, in case $A \not\sqsubseteq^{\max} B$, we can always find an adversary, modelled by a vulnerability function V, who prefers (wrt the max-case) interacting with A that with B. Such a function is discussed in Section 5.

Finally, we resolve the question of how $\sqsubseteq^{\max}$ and $\sqsubseteq^{\text{avg}}$ are related.

Theorem 3. *$\sqsubseteq^{\text{avg}}$ is strictly stronger than $\sqsubseteq^{\max}$.*

This result might appear counter-intuitive at first; one might expect $A \sqsubseteq^{\max} B$ to imply $A \sqsubseteq^{\text{avg}} B$. To understand why it does not, note that the former only requires that, for each output y_B of B, there exists some output of y_A that is at least as vulnerable, regardless of how likely y_A and y_B are to happen (this is max-case, after all). We illustrate this with the following example.

Example 6. *Consider the following systems:*

$$\begin{array}{c|cccc} A & y_1 & y_2 & y_3 & y_4 \\ \hline x_1 & 3/4 & 0 & 1/4 & 0 \\ x_2 & 3/4 & 1/4 & 0 & 0 \\ x_3 & 0 & 1/4 & 1/4 & 1/2 \end{array} \qquad \begin{array}{c|ccc} B & y_1 & y_2 & y_3 \\ \hline x_1 & 1/2 & 0 & 1/2 \\ x_2 & 1/2 & 1/2 & 0 \\ x_3 & 0 & 1/2 & 1/2 \end{array} \tag{14}$$

Under the uniform prior, the y_1, y_2, y_3 posteriors for both channels are the same, namely $(1/2, 1/2, 0)$, $(0, 1/2, 1/2)$ and $(1/2, 0, 1/2)$, respectively. Thus, the knowledge that can be obtained by each output of B can be also obtained by some output of A (albeit with a different probability). Hence, from Theorem A1, we get that $A \sqsubseteq^{\max} B$. However, we can check (see Section 5.1) that B cannot be obtained by post-processing A, that is, $A \not\sqsubseteq^{\text{avg}} B$.

The other direction might also appear tricky: if B leaks no more than A in the average-case, it must also leak no more than B in the max-case. The quantification over all gain functions in the average-case is powerful enough to "detect" differences in max-case leakage. The above result also means that $\sqsubseteq^{\text{avg}}$ could be useful even if we are interested in the max-case, since it gives us $\sqsubseteq^{\max}$ for free.

4. Privacy-Based Refinement

Thus far, we have compared systems based on their (average-case or max-case) leakage. In this section, we turn our attention to the model of differential privacy and discuss new ways of ordering mechanisms based on that model.

4.1. Differential Privacy and d-Privacy

Differential privacy relies on the observation that some pairs of secrets *need to be indistinguishable* from the point of view of the adversary in order to provide some meaningful notion of privacy; for instance, databases differing in a single individual should not be distinguishable, otherwise the privacy of that individual is violated. At the same, other pairs of secrets can be allowed to be distinguishable in order to provide some utility; for instance, distinguishing databases differing in many individuals allows us to answer a statistical query about those individuals.

This idea can be formalized by a *distinguishability metric* d. (To be precise, d is an extended pseudo metric, namely one in which distinct secrets can have distance 0, and distance $+\infty$ is allowed.) Intuitively, $d(x, x')$ models how *distinguishable* we allow these secrets to be. A value 0 means that we require x and x' to be completely indistinguishable to the adversary, while $+\infty$ means that she can distinguish them completely.

In this context, a mechanism is simply a channel (the two terms will be used interchangeably), mapping secrets $\mathcal{X}$ to some observations $\mathcal{Y}$. Denote by $\mathbb{M}\mathcal{X}$ the set of all metrics on $\mathcal{X}$. Given $d \in \mathbb{M}\mathcal{X}$, we define d-privacy as follows:

Definition 3. *A channel C satisfies d-privacy iff*

$$C_{x,y} \leq e^{d(x,x')} C_{x',y} \qquad \text{for all } x, x' \in \mathcal{X}, y \in \mathcal{Y} \tag{15}$$

Intuitively, this definition requires that the closer x and x' are (as measured by d), the more similar (probabilistically) the output of the mechanism on these secrets should be.

Remark 1. *Note that the definition of d-privacy given in (1) is slightly different from the above one because of the presence of ε in the exponent. Indeed, it is common to scale d by a privacy parameter $\varepsilon \geq 0$, in which case d can be thought of as the "kind" and ε as the "amount" of privacy. In other words, the structure determined by d* *on the data specifies how we want to distinguish each pair of data, and ε specifies (uniformly) the degree of the distinction. Note that $\varepsilon \cdot d$ is itself a metric, so the two definitions are equivalent.*

Using a generic metric d in this definition allows us to express different scenarios, depending on the domain $\mathcal{X}$ on which the mechanism is applied and the choice of d. For instance, in the standard model of differential privacy, the mechanism is applied to a database x (i.e., $\mathcal{X}$ is the set of all databases), and produces some observation y (e.g., a number). The *Hamming* metric d_{H}—defined as the number of individuals in which x and x' differ—captures standard differential privacy.

4.2. Oblivious Mechanisms

In the case of an *oblivious* mechanism, a query $f \colon \mathcal{X} \to \mathcal{Y}$ is first applied to database x, and a noise mechanism H from $\mathcal{Y}$ to $\mathcal{Z}$ is applied to $y = f(x)$, producing an observation z. In this case, it is useful to study the privacy of H wrt some metric $d_{\mathcal{Y}}$ on $\mathcal{Y}$. Then, to reason about the $d_{\mathcal{X}}$-privacy of the whole mechanism $H \circ f$, we can first compute the *sensitivity* of f wrt $d_{\mathcal{X}}, d_{\mathcal{Y}}$:

$$\Delta^{f}_{d_{\mathcal{X}},d_{\mathcal{Y}}} = \max_{x,x'} \frac{d_{\mathcal{Y}}(f(x), f(x'))}{d_{\mathcal{X}}(x, x')} \tag{16}$$

and then use the following property [3]:

$$\text{If } H \text{ satisfies } d_{\mathcal{Y}}\text{-privacy, then } H \circ f \text{ satisfies } \Delta^{f}_{d_{\mathcal{X}},d_{\mathcal{Y}}} \cdot d_{\mathcal{X}}\text{-privacy}\,. \tag{17}$$

For instance, the geometric mechanism G^{ε} satisfies $\varepsilon \cdot d_{\mathrm{E}}$-privacy (where d_{E} denotes the Euclidean metric), hence it can be applied to any numeric query f: the resulting mechanism $G^{\varepsilon} \circ f$ satisfies $\Delta^{f}_{d_{\mathrm{H}},d_{\mathrm{E}}} \cdot \varepsilon$-differential privacy. The sensitivity wrt the Hamming and Euclidean metrics reduces to $\Delta^{f}_{d_{\mathrm{H}},d_{\mathrm{E}}} = \max_{x \sim x'} |f(x) - f(x')|$ where $x \sim x'$ denotes $d_{\mathrm{H}}(x, x') = 1$.

4.3. Applying Noise to the Data of a Single Individual

There are also scenarios in which a mechanism C is applied directly to the data of a single individual (that is, $\mathcal{X}$ is the set of possible values). For instance, in the *local model* of differential privacy [2], the value of each individual is obfuscated before sending them to an untrusted curator. In this case, C should satisfy d_{D}-privacy, where d_{D} is the discrete metric, since *any change* in the individual's value should have negligible effects.

Moreover, in the context of *location-based services*, a user might want to obfuscate his location before sending it to the service provider. In this context, it is natural to require that locations that are geographically close are indistinguishable, while far away ones are allowed to be distinguished (in order to provide the service). In other words, we wish to provide d_{E}-privacy, for the Euclidean metric on $\mathbb{R}^2$, called *geo-indistinguishability* in [17].

4.4. Comparing Mechanisms by Their "Smallest ε" (For Fixed d)

Scaling d by a privacy parameter ε allows us to turn d-privacy (for some fixed d) into a *quantitative "leakage" measure*, by associating each channel to the *smallest* ε by which we can scale d without violating privacy.

Definition 4. *The privacy-based leakage (wrt d) of a channel C is defined as*

$$\mathrm{Priv}_d(C) \quad := \quad \inf\{\varepsilon \geq 0 \mid C \textit{ satisfies } \varepsilon \cdot d\textit{-privacy}\} \tag{18}$$

Note that $\mathrm{Priv}_d(C) = +\infty$ iff there is no such ε; also $\mathrm{Priv}_d(C) \leq 1$ iff C satisfies d-privacy.

It is then natural to compare two mechanisms A and B based on their "smallest ε".

Definition 5. *Define* $A \sqsubseteq_d^{\mathrm{prv}} B$ *iff* $\mathrm{Priv}_d(A) \geq \mathrm{Priv}_d(B)$.

For instance, $A \sqsubseteq_{d_{\mathrm{H}}}^{\mathrm{prv}} B$ means that B satisfies standard differential privacy for ε at least as small as the one of A.

4.5. Privacy-Based Leakage and Refinement Orders

When discussing the average- and max-case leakage orders $\sqsubseteq_{\mathbb{G}}^{\mathrm{avg}}, \sqsubseteq_{\mathbb{Q}}^{\max}$, we obtained strong leakage guarantees by quantifying over *all vulnerability functions*. It is thus natural to investigate a similar quantification in the context of d-privacy. Namely, we define a stronger privacy-based "leakage" order, by comparing mechanisms not on a single metric d, but on *all metrics* simultaneously.

Definition 6. *The privacy-based leakage order is defined as* $A \sqsubseteq_{\mathbb{M}}^{\mathrm{prv}} B$ *iff* $A \sqsubseteq_d^{\mathrm{prv}} B$ *for all* $d \in \mathbb{M}\mathcal{X}$.

Similarly to the other leakage orders, the drawback of $\sqsubseteq_{\mathbb{M}}^{\mathrm{prv}}$ is that it quantifies over an uncountable family of metrics. As a consequence, our first goal is to characterize it as a property of the channel matrix alone, which would make it much easier to reason about or verify.

To do so, we start by recalling an alternative way of thinking about d-privacy. Consider the *multiplicative total variation* distance between probability distributions $\mu, \mu' \in \mathbb{D}\mathcal{Y}$, defined as:

$$\mathrm{tv}_{\otimes}(\mu, \mu') \quad := \quad \max_{y:\mathcal{Y}} \left| \ln \frac{\mu_y}{\mu'_y} \right| \tag{19}$$

If we think of C as a function $\mathcal{X} \to \mathbb{D}\mathcal{Y}$ (mapping every x to the distribution $C_{x,-}$), C satisfies d-privacy iff $\mathrm{tv}_{\otimes}(C_{x,-}, C_{x',-}) \leq d(x, x')$, in other words iff C is non-expansive (1-Lipschitz) wrt $\mathrm{tv}_{\otimes}, d$.

Then, we introduce the concept of the *distinguishability metric* $d_C \in \mathbb{M}\mathcal{X}$ *induced by* the channel C, defined as

$$d_C(x, x') \quad := \quad \mathrm{tv}_{\otimes}(C_{x,-}, C_{x',-}) \tag{20}$$

Intuitively, $d_C(x, x')$ expresses exactly how much the channel distinguishes (wrt $\mathrm{tv}_{\otimes}$) the secrets x, x'. It is easy to see that d_C is the *smallest metric* for which C is private; in other words, for any d:

$$C \text{ satisfies } d\text{-privacy} \qquad \text{iff} \qquad d \geq d_C \tag{21}$$

We can now give a refinement order on mechanisms, by comparing their corresponding induced metrics.

Definition 7. *The privacy-based refinement order is defined as* $A \sqsubseteq^{\mathrm{prv}} B$ *iff* $d_A \geq d_B$, *or equivalently iff* B *satisfies* d_A*-privacy.*

This achieves our goal of goal of characterizing $\sqsubseteq_{\mathbb{M}}^{\mathrm{prv}}$.

Proposition 1. *The orders* $\sqsubseteq_{\mathbb{M}}^{\mathrm{prv}}$ *and* $\sqsubseteq^{\mathrm{prv}}$ *coincide.*

We now turn our attention to the question of how these orders relate to each other.

Theorem 4. $\sqsubseteq^{\max}$ *is strictly stronger than* $\sqsubseteq^{\mathrm{prv}}$, *which is strictly stronger than* $\sqsubseteq_d^{\mathrm{prv}}$.

The fact that $\sqsubseteq^{\max}$ is stronger than $\sqsubseteq^{\mathrm{prv}}$ is due to the fact than Priv_d can be seen as a max-case information leakage, for a properly constructed vulnerability function V_d. This is discussed in detail in Section 4.7. This implication means that $\sqsubseteq^{\mathrm{avg}}, \sqsubseteq^{\max}$ can be useful even if we "only" care about d-privacy.

4.6. Application to Oblivious Mechanisms

We conclude the discussion on privacy-based refinement by showing the usefulness of our strong $\sqsubseteq^{\mathrm{prv}}$ order in the case of oblivious mechanisms.

Theorem 5. *Let $f : \mathcal{X} \to \mathcal{Y}$ be any query and A, B be two mechanisms on $\mathcal{Y}$. If $A \sqsubseteq^{\mathrm{prv}} B$, then $A \circ f \sqsubseteq^{\mathrm{prv}} B \circ f$.*

This means that replacing A by B is the context of an oblivious mechanism is always safe, regardless of the query (and its sensitivity) and regardless of the metric by which the privacy of the composed mechanism is evaluated.

Assume, for instance that we care about standard differential privacy, and we have properly constructed A such that $A \circ f$ satisfies ε-differential privacy for some ε. If we know that $A \sqsubseteq^{\mathrm{prv}} B$ (several such cases are discussed in Section 6), we can replace A by B without even knowing what f does. The mechanism $B \circ f$ is guaranteed to also satisfy ε-differential privacy.

Note also that the above theorem fails for the weaker order $\sqsubseteq^{\mathrm{prv}}_{d}$. Establishing $A \sqsubseteq^{\mathrm{prv}}_{d_{\mathcal{Y}}} B$ for some metric $d_{\mathcal{Y}} {:} \mathbb{M}\mathcal{Y}$ gives no guarantees that $A \circ f \sqsubseteq^{\mathrm{prv}}_{d_{\mathcal{X}}} B \circ f$ for some other metric of interest $d_{\mathcal{X}} {:} \mathbb{M}\mathcal{X}$. It is possible that replacing A by B in that case is not safe (one would need to re-examine the behavior of B, and possibly reconfigure it to the sensitivity of f).

Table 3 summarizes the relations between the various orderings.

Table 3. Comparison of leakage and refinement orders. All implications are strict.

Leakage Orders		Refinement Orders
$\sqsubseteq^{\mathrm{avg}}_{\mathbb{G}}$	$\Leftrightarrow$	$\sqsubseteq^{\mathrm{avg}}$
$\Downarrow$		$\Downarrow$
$\sqsubseteq^{\mathrm{max}}_{\mathbb{Q}}$	$\Leftrightarrow$	$\sqsubseteq^{\mathrm{max}}$
$\Downarrow$		$\Downarrow$
$\sqsubseteq^{\mathrm{prv}}_{\mathbb{M}}$	$\Leftrightarrow$	$\sqsubseteq^{\mathrm{prv}}$
$\Searrow$		$\Swarrow$
	$\sqsubseteq^{\mathrm{prv}}_{d}$	

4.7. Privacy as Max-Case Capacity

In this section we show that d-privacy can be expressed as a (max-case) information leakage. Note that this provides an alternative proof that $\sqsubseteq^{\mathrm{max}}$ is stronger than $\sqsubseteq^{\mathrm{prv}}$. We start this by defining a suitable vulnerability function:

Definition 8. *The d-vulnerability function V_d is defined as*

$$V_d(\pi) \quad := \quad \inf\{\varepsilon \geq 0 \mid \forall x, x' \in \mathcal{X}, \pi_x \leq e^{\varepsilon \cdot d(x,x')} \pi_{x'}\} \tag{22}$$

Note the difference between $V_d(\pi)$ (a vulnerability function on *distributions*) and $\mathrm{Priv}_d(C)$ (a “leakage” measure on *channels*).

A fundamental notion in QIF is that of *capacity*: the maximization of leakage over all priors. In turns out that, for V_d, the capacity-realizing prior is the uniform one. In the following, $\mathcal{L}^{+,\mathrm{max}}_{d}$ denotes the additive max-case d leakage, namely:

$$\mathcal{L}^{+,\mathrm{max}}_{d}(\pi, C) = V^{\mathrm{max}}_{d}[\pi, C] - V_d(\pi) \tag{23}$$

and $\mathcal{ML}^{+,\mathrm{max}}_{d}$ denotes the additive max-case d-capacity, namely:

$$\mathcal{ML}^{+,\mathrm{max}}_{d}(C) = \max_{\pi} \mathcal{L}^{+,\mathrm{max}}_{d}(\pi, C) \tag{24}$$

Theorem 6. *$\mathcal{ML}_d^{+,\max}$ is always achieved on a uniform prior π^u. Namely*

$$\max_\pi \mathcal{L}_d^{+,\max}(\pi, C) \quad = \quad \mathcal{L}_d^{+,\max}(\pi^u, C) \quad = \quad V_d^{\max}[\pi^u, C] \tag{25}$$

This finally brings us to our goal of expressing Priv_d in terms of information leakage (for a proper vulnerability function).

Theorem 7. *[DP as max-case capacity] C satisfies $\varepsilon \cdot d$-privacy iff $\mathcal{ML}_d^{+,\max}(C) \le \varepsilon$. In other words: $\mathcal{ML}_d^{+,\max}(C) = \mathrm{Priv}_d(C)$.*

5. Verifying the Refinement Orders

We now turn our attention to the problem of checking whether the various orders hold, given two *explicit representations* of channels A and B (in terms of their matrices). We show that, for all orders, this question can be answered in time polynomial in the size of the matrices. Moreover, when one of the order fails, we discuss how to obtain a *counter-example* (e.g., a gain function g or a vulnerability function V), demonstrating this fact. All the methods discussed in the section have been implemented in a publicly available library, and have been used in the experimental results of Section 6.

5.1. Average-Case Refinement

Verifying $A \sqsubseteq^{\mathrm{avg}} B$ can be done in polynomial time (in the size of A, B) by solving the system of equations $AR = B$, with variables R, under the linear constraints that R is a channel matrix (non-negative and rows sum up to 1). However, if the system has no solution (i.e., $A \not\sqsubseteq^{\mathrm{avg}} B$), this method does not provide us with a counter-example gain function g.

We now show that there is an alternative efficient method: define $C^\uparrow = \{CR \mid R \text{ is a channel}\}$, the set of all channels obtainable by post-processing C. The idea is to compute the projection of B on $A^\uparrow$. Clearly, the projection is B itself iff $A \sqsubseteq^{\mathrm{avg}} B$; otherwise, the projection can be directly used to construct a counter-example g.

Theorem 8. *Let B^* be the projection of B on $A^\uparrow$.*

1. *If $B = B^*$, then $A \sqsubseteq^{\mathrm{avg}} B$.*
2. *Otherwise, let $G = B - B^*$. The gain function $g(w, x) = G_{x,w}$ provides a counter-example to $A \sqsubseteq^{\mathrm{avg}} B$, which is $V_g(\pi^u, A) < V_g(\pi^u, B)$, for uniform π^u.*

Since $\|x - y\|_2^2 = x^T x - 2x^T y + y^T y$, the projection of y to a convex set can be written as $\min_x x^T x - 2x^T y$ for $Ax \le b$. This is a quadratic program with Q being the identity matrix, which is positive definite, hence it can be solved in polynomial time.

Note that the proof that $\sqsubseteq_{\mathbb{G}}^{\mathrm{avg}}$ is stronger than $\sqsubseteq^{\mathrm{avg}}$ (the "coriaceous" theorem of [7]) uses the hyperplane-separation theorem to show the existence of a counter example g in case $A \not\sqsubseteq^{\mathrm{avg}} B$. The above technique essentially computes such a separating hyperplane.

5.2. Max-Case Refinement

Similarly to $\sqsubseteq^{\mathrm{avg}}$, we can verify $A \sqsubseteq^{\max} B$ directly using its definition, by solving the system $R\tilde{A} = \tilde{B}$ under the constraint that R is a channel.

In contrast to $\sqsubseteq^{\mathrm{avg}}$, when $A \not\sqsubseteq^{\max} B$, the proof of Theorem 2 directly gives us a counter-example:

$$V(\sigma) \quad := \quad \min_{\sigma' : S} \|\sigma - \sigma'\|_2 \tag{26}$$

where $S = \mathit{ch\,supp}\,[\pi, A]$ and π is any full-support prior. For this vulnerability function, it holds that $V^{\max}(\pi, A) < V^{\max}(\pi, B)$.

5.3. Privacy-Based Refinement

The $\sqsubseteq^{\text{prv}}$ order can be verified directly from its definition, by checking that $d_A \geq d_B$. This can be done in time $O(|\mathcal{X}|^2|\mathcal{Y}|)$, by computing $\text{tv}_{\otimes}(C_{x,-}, C_{x',-})$ for each pair of secrets. If $A \not\sqsubseteq^{\text{prv}} B$, then $d = d_B$ provides an immediate counter-example metric, since B satisfies d_B-privacy, but A does not.

6. Application: Comparing DP Mechanisms

In differential privacy, it is common to compare the privacy guarantees provided by different mechanisms by 'comparing the epsilons'. However, it is interesting to ask to what extent ε-equivalent mechanisms are comparable wrt the other leakage measures defined here—or we might want to know whether reducing ε in a mechanism also corresponds to a *refinement* of it. This could be useful if, for example, it is important to understand the privacy properties of a mechanism with respect to *any* max-case leakage measure, and not just the DP measure given by ε.

Since the ε-based order given by $\sqsubseteq_d^{\text{prv}}$ is (strictly) the weakest of the orders considered here, it cannot be the case that we *always* get a refinement (wrt other orders). However, it may be true that, for particular *families* of mechanisms, some (or all) of the refinement orders hold.

We investigate three families of mechanisms commonly used in DP or LDP: *geometric*, *exponential* and *randomized response* mechanisms.

6.1. Preliminaries

We define each family of mechanisms in terms of their channel construction. We assume that mechanisms operate on a set of inputs (denoted by $\mathcal{X}$) and produce a set of outputs (denoted $\mathcal{Y}$). In this sense, our mechanisms can be seen as oblivious (as in standard DP) or as LDP mechanisms. (We use the term 'mechanism' in either sense). We denote by M^ε a mechanism parametrized by ε, where ε is defined to be the same as $\text{Priv}_d(M)$. For the purposes of comparison, we make sure that we use the best possible ε for each mechanism. In order to compare mechanisms, we restrict our input and output domains of interest to sequences of non-negative integers. We assume $\mathcal{X}, \mathcal{Y}$ are finite unless specified. In addition, as we are operating in the framework of d-privacy, it is necessary to provide an appropriate metric defined over $\mathcal{X}$; here, it makes sense to use the Euclidean distance metric d_{E}.

Definition 9. *A geometric mechanism is a channel $(\mathcal{X}, \mathbb{Z}, G^\varepsilon)$, parametrized by $\varepsilon \geq 0$ constructed as follows:*

$$G^\varepsilon_{x,y} = \frac{(1-\alpha)\cdot \alpha^{d_E(x,y)}}{1+\alpha} \qquad \text{for all } x \in \mathcal{X}, y \in \mathbb{Z} \tag{27}$$

where $\alpha = e^{-\varepsilon}$ and $d_{\text{E}}(x,y) = ||x-y||$. Such a mechanism satisfies $\varepsilon \cdot d_{\text{E}}$-privacy.

In practice, the truncated geometric mechanism is preferred to the infinite geometric. We define the truncated geometric mechanism as follows.

Definition 10. *A truncated geometric mechanism is a channel $(\mathcal{X}, \mathcal{Y}, TG^\varepsilon)$, parametrized by $\varepsilon \geq 0$ with $\mathcal{X} \subseteq \mathcal{Y}$ constructed as follows:*

$$TG^\varepsilon_{x,y} = \frac{(1-\alpha)\cdot \alpha^{d_E(x,y)}}{1+\alpha} \qquad \text{for all } y \neq \min\mathcal{Y}, \max\mathcal{Y} \tag{28}$$

$$TG^\varepsilon_{x,y} = \frac{\alpha^{d_E(x,y)}}{1+\alpha} \qquad \text{for } y = \min\mathcal{Y}, \max\mathcal{Y} \tag{29}$$

where $\alpha = e^{-\varepsilon}$ and $d_{\text{E}}(x,y) = ||x-y||$. Such a mechanism satisfies $\varepsilon \cdot d_{\text{E}}$-privacy.

It is also possible to define the 'over-truncated' geometric mechanism whose input space is not entirely included in the output space.

Definition 11. *An over-truncated geometric mechanism is a channel $(\mathcal{X}, \mathcal{Y}, OTG^{\varepsilon})$, parametrized by $\varepsilon \geq 0$ with $\mathcal{X} \not\subseteq \mathcal{Y}$ constructed as follows:*

1. *Start with the truncated geometric mechanism $(\mathcal{X}, \mathcal{X} \cup \mathcal{Y}, TG^{\varepsilon})$.*
2. *Sum up the columns at each end until the output domain is reached.*

Such a mechanism satisfies $\varepsilon \cdot d_E$-privacy.

For example, the set of inputs to an over-truncated geometric mechanism could be integers in the range $[0 \ldots 100]$, but the output space may have a range of $[0 \ldots 50]$ or perhaps $[-50 \ldots 50]$. In either of these cases, the mechanism has to 'over-truncate' the inputs to accommodate the output space.

We remark that we do not consider the over-truncated mechanism a particularly useful mechanism in practice. However, we provide results on this mechanism for completeness since its construction is possible, if unusual.

Definition 12. *An exponential mechanism is a channel $(\mathcal{X}, \mathcal{Y}, E^{\alpha})$, parametrized by $\varepsilon \geq 0$ constructed as follows:*

$$E^{\alpha}_{x,y} = \lambda_x \cdot e^{-\frac{\varepsilon}{2} d_E(x,y)} \qquad \textit{for all } x \in \mathcal{X}, y \in \mathcal{Y} \tag{30}$$

where λ_x are normalizing constants ensuring $\sum_y E^{\alpha}_{x,y} = 1$. Such a mechanism satisfies $\alpha \cdot d_E$-privacy where $\alpha \geq \frac{\varepsilon}{2}$ (which can be calculated exactly from the channel construction).

Note that the construction presented in Definition 12 uses the Euclidean distance metric since we only consider integer domains. The general construction of the exponential mechanism uses arbitrary domains and arbitrary metrics. Note that its parameter ε does not correspond to the true (best-case) ε-DP guarantee that it provides. We will denote by E^{ε} the exponential mechanism with 'true' privacy parameter ε rather than the reported one, as our intention is to capture the privacy guarantee provided by the channel in order to make reasonable comparisons.

Definition 13. *A randomized response mechanism is a channel $(\mathcal{X}, \mathcal{Y}, R^{\varepsilon})$, parametrized by $\varepsilon \geq 0$ constructed as follows:*

$$R^{\varepsilon}_{x,y} = \frac{e^{\varepsilon(1-d_D(x,y))}}{e^{\varepsilon} + n} \qquad \textit{for all } x, y \in \mathcal{Y} \tag{31}$$

$$R^{\varepsilon}_{x,y} = \frac{1}{n+1} \qquad \textit{for all } x \notin \mathcal{Y} \tag{32}$$

where $n = \|\mathcal{Y}\| - 1$ and d_D is the discrete metric (that is, $d_D(x,x) = 0$ and $d_D(x,y) = 1$ for $x \in \mathcal{Y}, x \neq y$). Such a mechanism satisfies $\varepsilon \cdot d_D$-privacy.

We note that the randomized response mechanism also satisfies $\varepsilon \cdot d_E$-privacy.

Intuitively, the randomized response mechanism returns the true answer with high probability and all other responses with equal probability. In the case where the input x lies outside $\mathcal{Y}$ (that is, in 'over-truncated' mechanisms), all of the outputs (corresponding to the outlying inputs) have equal probability.

Example 7. *The following are examples of each of the mechanisms described above, represented as channel matrices. For this example, we set $\varepsilon = \log(2)$ for the geometric and randomized response mechanisms, while, for the exponential mechanism, we use $\varepsilon = \log(4)$.*

$$
\begin{array}{c|ccc}
TG & x_1 & x_2 & x_3 \\ \hline
x_1 & 2/3 & 1/6 & 1/6 \\
x_2 & 1/3 & 1/3 & 1/3 \\
x_3 & 1/6 & 1/6 & 2/3
\end{array}
\qquad
\begin{array}{c|cc}
OTG & x_1 & x_2 \\ \hline
x_1 & 2/3 & 1/3 \\
x_2 & 1/3 & 2/3 \\
x_3 & 1/6 & 5/6
\end{array}
$$

$$
\begin{array}{c|ccc}
E & x_1 & x_2 & x_3 \\ \hline
x_1 & 4/7 & 2/7 & 1/7 \\
x_2 & 1/4 & 1/2 & 1/4 \\
x_3 & 1/7 & 2/7 & 4/7
\end{array}
\qquad
\begin{array}{c|ccc}
R & x_1 & x_2 & x_3 \\ \hline
x_1 & 1/2 & 1/4 & 1/4 \\
x_2 & 1/4 & 1/2 & 1/4 \\
x_3 & 1/4 & 1/4 & 1/2
\end{array}
\tag{33}
$$

Note that the exponential mechanism here actually satisfies $\log(\frac{16}{7}) \cdot d_E$*-privacy even though it is specified by* $\varepsilon = \log(4)$.

We now have three families of mechanisms which we can characterize by channels, and which satisfy $\varepsilon \cdot d_E$-privacy. For the remainder of this section, we will refer only to the ε parameter and take d_E as given, as we wish to understand the effect of changing ε (for a fixed metric) on the various leakage measures.

6.2. Refinement Order within Families of Mechanisms

We first ask which refinement orders hold within a family of mechanisms. That is, when does reducing ε for a particular mechanism produce a refinement? Since we have the convenient order $\sqsubseteq^{\text{avg}} \subset \sqsubseteq^{\max} \subset \sqsubseteq^{\text{prv}}$, it is useful to first check if $\sqsubseteq^{\text{avg}}$ holds as we get the other refinements 'for free'.

For the (infinite) geometric mechanism, we have the following result.

Theorem 9. *Let* $G^{\varepsilon}, G^{\varepsilon'}$ *be geometric mechanisms. Then,* $G^{\varepsilon} \sqsubseteq^{\text{avg}} G^{\varepsilon'}$ *iff* $\varepsilon \geq \varepsilon'$*. That is, decreasing* ε *produces a refinement of the mechanism.*

This means that reducing ε in an infinite geometric mechanism is safe against *any* adversary that can be modelled using, for example, max-case or average-case vulnerabilities.

For the truncated geometric mechanism, we get the same result.

Theorem 10. *Let* $TG^{\varepsilon}, TG^{\varepsilon'}$ *be truncated geometric mechanisms. Then,* $TG^{\varepsilon} \sqsubseteq^{\text{avg}} TG^{\varepsilon'}$ *iff* $\varepsilon \geq \varepsilon'$*. That is, decreasing* ε *produces a refinement of the mechanism.*

However, the over-truncated geometric mechanism does not behave so well.

Theorem 11. *Let* $OTG^{\varepsilon}, OTG^{\varepsilon'}$ *be over-truncated geometric mechanisms. Then,* $OTG^{\varepsilon} \not\sqsubseteq^{\text{avg}} OTG^{\varepsilon'}$ *for any* $\varepsilon \neq \varepsilon'$*. That is, decreasing* ε *does* ***not*** *produce a refinement.*

We can think of this last class of geometrics as 'skinny' mechanisms, that is, corresponding to a channel with a smaller output space than input space.

Intuitively, this theorem means that we can *always* find some (average-case) adversary who prefers the over-truncated geometric mechanism with the smaller ε.

We remark that the gain function we found can be easily calculated by treating the columns of channel A as vectors, and finding a vector orthogonal to both of these. This follows from the results in Section 5.1. Since the columns of A cannot span the space $\mathbb{R}^3$, it is always possible to find such a vector, and, when this vector is not orthogonal to the 'column space' of B, it can be used to construct a gain function preferring B to A.

Even though the $\sqsubseteq^{\text{avg}}$ refinement does not hold, we can check whether the other refinements are satisfied.

Theorem 12. *Let OTG^{ε} be an over-truncated geometric mechanism. Then, reducing ε does **not** produce a $\sqsubseteq^{\max}$ refinement; however, it **does** produce a $\sqsubseteq^{\mathrm{prv}}$ refinement.*

This means that, although a smaller ε does not provide safety against all max-case adversaries, it *does* produce a safer mechanism wrt d-privacy for *any* choice of metric we like.

Intuitively, the $\sqsubseteq^{\mathrm{prv}}$ order relates mechanisms based on how they distinguish *inputs*. Specifically, if $A \sqsubseteq^{\mathrm{prv}} B$, then, for any pair of inputs x, x', the corresponding output distributions are 'further apart' in channel A than in channel B, and thus the inputs are more distinguishable using channel A. When $\sqsubseteq^{\mathrm{prv}}$ fails to hold, it means that there are some inputs in A which are more distinguishable than in B, and vice versa. This means an adversary who is interested in distinguishing some particular pair of inputs would prefer one mechanism to the other.

We now consider the exponential mechanism. In this case, we do not have a theoretical result, but, experimentally, it appears that the exponential mechanism respects refinement, so we present the following conjecture.

Conjecture 1. *Let E^{ε} be an exponential mechanism. Then, decreasing ε in E produces a refinement. That is, $E^{\varepsilon} \sqsubseteq^{\mathrm{avg}} E^{\varepsilon'}$ iff $\varepsilon \geq \varepsilon'$.*

Finally, we consider the randomized response mechanism.

Theorem 13. *Let R^{ε} be a randomized response mechanism. Then, decreasing ε in R produces a refinement. That is, $R^{\varepsilon} \sqsubseteq^{\mathrm{avg}} R^{\varepsilon'}$ iff $\varepsilon \geq \varepsilon'$.*

In conclusion, we can say that, in general, the usual DP families of mechanisms are 'well-behaved' wrt all of the refinement orders. This means that it is safe (wrt *any* adversary we model here) to replace a mechanism from a particular family with another mechanism from the *same* family with a lower ε.

6.3. Refinement Order between Families of Mechanisms

Now, we explore whether it is possible to compare mechanisms from different families. We first ask: can we compare mechanisms which have the same ε? We assume that the input and output domains are the same, and the intention is to decide whether to replace one mechanism with another.

Theorem 14. *Let R be a randomized response mechanism, E an exponential mechanism and TG a truncated geometric mechanism. Then, $TG^{\varepsilon} \sqsubseteq^{\mathrm{prv}} R^{\varepsilon}$ and $TG^{\varepsilon} \sqsubseteq^{\mathrm{prv}} E^{\varepsilon}$. However, $\sqsubseteq^{\mathrm{prv}}$ does not hold between E^{ε} and R^{ε}.*

Proof. We present a counter-example to show $E^{\varepsilon} \not\sqsubseteq^{\mathrm{prv}} R^{\varepsilon}$ and $R^{\varepsilon} \not\sqsubseteq^{\mathrm{prv}} E^{\varepsilon}$. The remainder of the proof is in Appendix D.

Consider the following channels:

$$
\begin{array}{c|cccc}
A & x_1 & x_2 & x_3 & x_4 \\
\hline
x_1 & 8/15 & 4/15 & 2/15 & 1/15 \\
x_2 & 2/9 & 4/9 & 2/9 & 1/9 \\
x_3 & 1/9 & 2/9 & 4/9 & 2/9 \\
x_4 & 1/15 & 2/15 & 4/15 & 8/15
\end{array}
\qquad
\begin{array}{c|cccc}
B & x_1 & x_2 & x_3 & x_4 \\
\hline
x_1 & 4/9 & 5/27 & 5/27 & 5/27 \\
x_2 & 5/27 & 4/9 & 5/27 & 5/27 \\
x_3 & 5/27 & 5/27 & 4/9 & 5/27 \\
x_4 & 5/27 & 5/27 & 5/27 & 4/9
\end{array}
\tag{34}
$$

Channel A represents an exponential mechanism and channel B a randomized response mechanism. Both have (true) ε of $\log(12/5)$. (Channel A was generated using $\varepsilon = \log(4)$. However, as noted earlier, this corresponds to a lower *true* ε.) However, $d_A(x_1, x_3) > d_B(x_1, x_3)$ and $d_A(x_2, x_3) < d_B(x_2, x_3)$. Thus, A does not satisfy d_B-privacy, nor does B satisfy d_A-privacy. □

Intuitively, the randomized response mechanism maintains the same (ε) distinguishability level between inputs, whereas the exponential mechanism causes some inputs to be *less* distinguishable than others. This means that, for the same (true) ε, an adversary who is interested in certain inputs could learn more from the randomized response than the exponential. In the above counter-example, points x_2, x_3 in the exponential mechanism of channel A are *less* distinguishable than the corresponding points in the randomized response mechanism B.

As an example, let's say the mechanisms are to be used in geo-location privacy and the inputs represent adjacent locations (such as addresses along a street). Then, an adversary (your boss) may be interested in how far you are from work, and therefore wants to be able to distinguish between points distant from x_1 (your office) and points within the vicinity of your office, without requiring your precise location. Your boss chooses channel A as the most informative. However, another adversary (your suspicious partner) is more concerned about where exactly you are, and is particularly interested in distinguishing between your expected position (x_2, the boulangerie) versus your suspected position (x_3, the brothel). Your partner chooses channel B as the most informative.

Regarding the other refinements, we find (experimentally) that in general they do not hold between families of mechanisms. (Recall that we only need to produce a single counter-example to show that a refinement doesn't hold, and this can be done using the methods presented in Section 5.)

We next check what happens when we compare mechanisms with *different* epsilons. We note the following.

Theorem 15. *For any (truncated geometric, randomized response, exponential) mechanisms $M_1^{\varepsilon_1}, M_2^{\varepsilon_2}$, if $M_1^{\varepsilon_1} \sqsubseteq M_2^{\varepsilon_2}$ for any of our refinements ($\sqsubseteq^{\text{avg}}, \sqsubseteq^{\max}, \sqsubseteq^{\text{prv}}$), then $M_1^{\varepsilon_1} \sqsubseteq M_2^{\varepsilon_2'}$ for $\varepsilon_2' < \varepsilon_2$.*

Proof. This follows directly from transitivity of the refinement relations, and our results on refinement with families of mechanisms. (We recall however that our result for the exponential mechanism is only a conjecture.) □

This tells us that, once we have a refinement between mechanisms, it continues to hold for reduced ε in the refining mechanism.

Corollary 1. *Let G, TG, R, E be the geometric, truncated geometric, randomized response and exponential mechanisms respectively. Then, for all $\varepsilon' \leq \varepsilon$, we have that $TG^{\varepsilon} \sqsubseteq^{\text{prv}} R^{\varepsilon'}$, $TG^{\varepsilon} \sqsubseteq^{\text{prv}} E^{\varepsilon'}$, $G^{\varepsilon} \sqsubseteq^{\text{prv}} R^{\varepsilon'}$ and $G^{\varepsilon} \sqsubseteq^{\text{prv}} E^{\varepsilon'}$.*

Thus, it is safe to 'compare epsilons' wrt $\sqsubseteq^{\text{prv}}$ if we want to replace a geometric mechanism with either a randomized response or exponential mechanism. (As with the previous theorem, note that the results for the exponential mechanism are stated as conjecture only, and this conjecture is assumed in the statement of this corollary.) What this means is that if, for example, we have a geometric mechanism TG that operates on databases with distance measured using the Hamming metric d_{H} and satisfying $\varepsilon \cdot d_{\text{H}}$-privacy, then any randomized response mechanism R parametrized by $\varepsilon' \leq \varepsilon$ will also satisfy $\varepsilon \cdot d_{\text{H}}$-privacy. Moreover, if we decide we'd rather use the Manhattan metric d_{M} to measure distance between the databases, then we only need to check that TG also satisfies $\varepsilon \cdot d_{\text{M}}$-privacy, as this implies that R will too.

The following Tables 4–6 summarize the refinement relations with respect to the various families of mechanisms.

Table 4. The refinements respected by families of mechanisms for decreasing ε. We recall that the results in the grey cells are based on Conjecture 1.

	Are These Valid for Decreasing ε?		
Mechanism	$\sqsubseteq^{\mathbf{avg}}$	$\sqsubseteq^{\mathbf{max}}$	$\sqsubseteq^{\mathbf{prv}}$
Geometric	Y	Y	Y
Truncated Geometric	Y	Y	Y
Over-Truncated Geometric	N	N	Y
Exponential	Y	Y	Y
Randomized Response	Y	Y	Y

Table 5. Comparing different families of mechanisms with respect to the different refinements under the same ε.

Refinements across Families with Same ε		
$TG \not\sqsubseteq^{\text{avg}} R$	$TG \not\sqsubseteq^{\text{max}} R$	$TG \sqsubseteq^{\text{prv}} R$
$R \not\sqsubseteq^{\text{avg}} TG$	$R \not\sqsubseteq^{\text{max}} TG$	$R \not\sqsubseteq^{\text{prv}} TG$
$TG \not\sqsubseteq^{\text{avg}} E$	$TG \not\sqsubseteq^{\text{max}} E$	$TG \sqsubseteq^{\text{prv}} E$
$E \not\sqsubseteq^{\text{avg}} TG$	$E \not\sqsubseteq^{\text{max}} TG$	$E \not\sqsubseteq^{\text{prv}} TG$
$G \not\sqsubseteq^{\text{avg}} R$	$G \not\sqsubseteq^{\text{max}} R$	$G \sqsubseteq^{\text{prv}} R$
$R \not\sqsubseteq^{\text{avg}} G$	$R \not\sqsubseteq^{\text{max}} G$	$R \not\sqsubseteq^{\text{prv}} G$
$G \not\sqsubseteq^{\text{avg}} E$	$G \not\sqsubseteq^{\text{max}} E$	$G \sqsubseteq^{\text{prv}} E$
$E \not\sqsubseteq^{\text{avg}} G$	$E \not\sqsubseteq^{\text{max}} G$	$E \not\sqsubseteq^{\text{prv}} G$
$R \not\sqsubseteq^{\text{avg}} E$	$R \not\sqsubseteq^{\text{max}} E$	$R \not\sqsubseteq^{\text{prv}} E$
$E \not\sqsubseteq^{\text{avg}} R$	$E \not\sqsubseteq^{\text{max}} R$	$E \not\sqsubseteq^{\text{prv}} R$

Table 6. Comparing different families of mechanisms with differing ε. We recall that the results in the grey cells are based on Conjecture 1.

Comparison of Refinements with $\varepsilon_1 > \varepsilon_2$.		
$TG^{\varepsilon_1} \not\sqsubseteq^{\text{avg}} R^{\varepsilon_2}$	$TG^{\varepsilon_1} \not\sqsubseteq^{\text{max}} R^{\varepsilon_2}$	$TG^{\varepsilon_1} \sqsubseteq^{\text{prv}} R^{\varepsilon_2}$
$R^{\varepsilon_1} \not\sqsubseteq^{\text{avg}} TG^{\varepsilon_2}$	$R^{\varepsilon_1} \not\sqsubseteq^{\text{max}} TG^{\varepsilon_2}$	$R^{\varepsilon_1} \not\sqsubseteq^{\text{prv}} TG^{\varepsilon_2}$
$TG^{\varepsilon_1} \not\sqsubseteq^{\text{avg}} E$	$TG^{\varepsilon_1} \not\sqsubseteq^{\text{max}} E^{\varepsilon_2}$	$TG^{\varepsilon_1} \sqsubseteq^{\text{prv}} E^{\varepsilon_2}$
$E^{\varepsilon_1} \not\sqsubseteq^{\text{avg}} TG$	$E^{\varepsilon_1} \not\sqsubseteq^{\text{max}} TG^{\varepsilon_2}$	$E^{\varepsilon_1} \not\sqsubseteq^{\text{prv}} TG^{\varepsilon_2}$
$G^{\varepsilon_1} \not\sqsubseteq^{\text{avg}} R^{\varepsilon_2}$	$G^{\varepsilon_1} \not\sqsubseteq^{\text{max}} R^{\varepsilon_2}$	$G^{\varepsilon_1} \sqsubseteq^{\text{prv}} R^{\varepsilon_2}$
$R^{\varepsilon_1} \not\sqsubseteq^{\text{avg}} G^{\varepsilon_2}$	$R^{\varepsilon_1} \not\sqsubseteq^{\text{max}} G^{\varepsilon_2}$	$R^{\varepsilon_1} \not\sqsubseteq^{\text{prv}} G^{\varepsilon_2}$
$G^{\varepsilon_1} \not\sqsubseteq^{\text{avg}} E^{\varepsilon_2}$	$G^{\varepsilon_1} \not\sqsubseteq^{\text{max}} E^{\varepsilon_2}$	$G^{\varepsilon_1} \sqsubseteq^{\text{prv}} E^{\varepsilon_2}$
$E^{\varepsilon_1} \not\sqsubseteq^{\text{avg}} G^{\varepsilon_2}$	$E^{\varepsilon_1} \not\sqsubseteq^{\text{max}} G^{\varepsilon_2}$	$E^{\varepsilon_1} \not\sqsubseteq^{\text{prv}} G^{\varepsilon_2}$
$R^{\varepsilon_1} \not\sqsubseteq^{\text{avg}} E^{\varepsilon_2}$	$R^{\varepsilon_1} \not\sqsubseteq^{\text{max}} E^{\varepsilon_2}$	$R^{\varepsilon_1} \not\sqsubseteq^{\text{prv}} E^{\varepsilon_2}$
$E^{\varepsilon_1} \not\sqsubseteq^{\text{avg}} R^{\varepsilon_2}$	$E^{\varepsilon_1} \not\sqsubseteq^{\text{max}} R^{\varepsilon_2}$	$E^{\varepsilon_1} \not\sqsubseteq^{\text{prv}} R^{\varepsilon_2}$

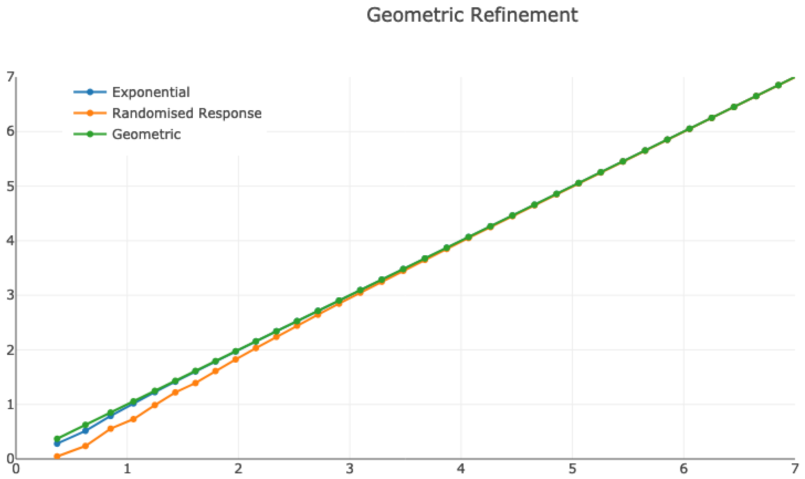
6.4. Asymptotic Behavior

We now consider the behavior of the relations when ε approximates 0, which represents the absence of leakage. We start with the following result:

Theorem 16. *Every (truncated geometric, randomized response, exponential) mechanism is 'the safest possible mechanism' when parametrized by $\varepsilon = 0$. That is, $L^{\varepsilon} \sqsubseteq^{\mathrm{avg}} M^0$ for all mechanisms L, M (possibly from different families) and $\varepsilon > 0$.*

While this result may be unsurprising, it means that we know that refinement must *eventually* occur when we reduce ε. It is interesting then to ask just *when* this refinement occurs. We examine this question experimentally by considering different mechanisms and investigating for which values of ε average-case refinement holds. For simplicity of presentation, we show results for 5×5 matrices, noting that we observed similar results for experiments across different matrix dimensions, at least wrt the coarse-grained comparison of plots that we do here. The results are plotted in Figure 3.

The plots show the relationship between ε_1 (x-axis) and ε_2 (y-axis) where ε_1 parametrizes the mechanism being refined and ε_2 parametrizes the refining mechanism. For example, the blue line on the top graph represents $TG^{\varepsilon_1} \sqsubseteq^{\mathrm{avg}} E^{\varepsilon_2}$. We fix ε_1 and ask for what value of ε_2 do we get a $\sqsubseteq^{\mathrm{avg}}$ refinement. Notice that the line $\varepsilon_1 = \varepsilon_2$ corresponds to the same mechanism in both axes (since every mechanism refines itself).



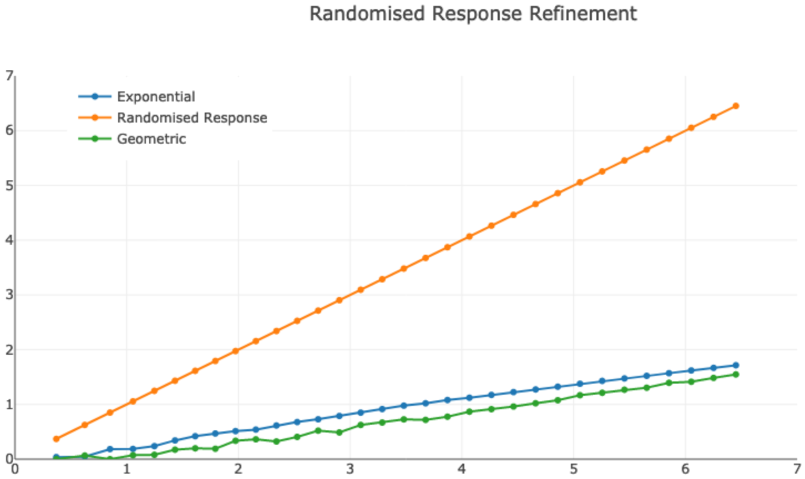


Figure 3. *Cont.*

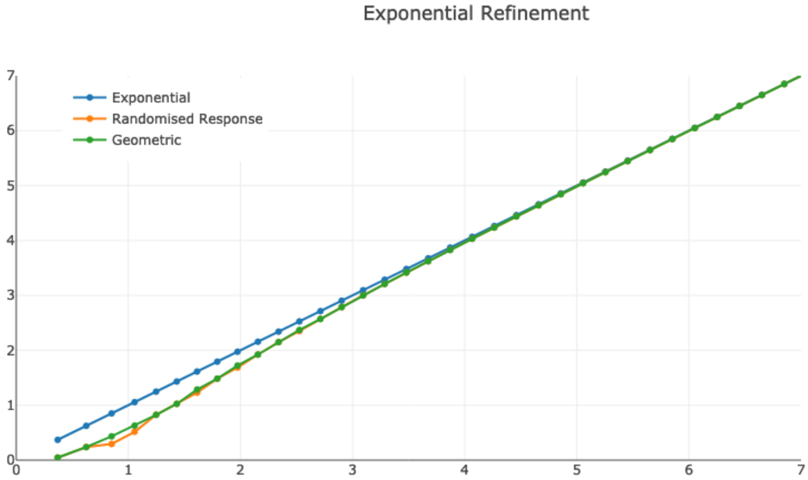


Figure 3. Refinement of mechanisms under $\sqsubseteq^{\text{avg}}$ for 5×5 channels. The x-axis represents the ε on the LHS of the relation, and the y-axis represents the one on the RHS. The top graph represents refinement of the truncated geometric mechanism (that is, $TG \sqsubseteq^{\text{avg}}$), the middle graph is refinement of randomized response ($R \sqsubseteq^{\text{avg}}$), and the bottom graph is refinement of the exponential mechanism ($E \sqsubseteq^{\text{avg}}$).

We can see that refining the randomized response mechanism requires much smaller values of epsilon in the other mechanisms. For example, from the middle graph, we can see that $R^4 \sqsubseteq^{\text{avg}} TG^1$ (approximately) whereas, from the top graph, we have $TG^4 \sqsubseteq^{\text{avg}} R^4$. This means that the randomized response mechanism is very 'safe' against average-case adversaries compared with the other mechanisms, as it is much more 'difficult' to refine than the other mechanisms.

We also notice that, for 'large' values of ε_1, the exponential and geometric mechanisms refine each other for approximately the same ε_2 values. This suggests that, for these values, the epsilons are comparable (that is, the mechanisms are equally 'safe' for similar values of ε). However, smaller values of ε_1 require a (relatively) large reduction in ε_2 to obtain a refinement.

6.5. Discussion

At the beginning of this section, we asked whether it is safe to compare differential privacy mechanisms by 'comparing the epsilons'. We found that it is safe to compare epsilons within families of mechanisms (except in the unusual case of the over-truncated geometric mechanism). However, when comparing different mechanisms, it is *not* safe to just compare the epsilons, since none of the refinements hold in general. Once a 'safe' pair of epsilons has been calculated, then reducing epsilon in the refining mechanism is always safe. However, computing safe epsilons relies on the ability to construct a channel representation, which may not always be feasible.

7. Lattice Properties

The orders $\sqsubseteq^{\text{avg}}, \sqsubseteq^{\text{max}}$ and $\sqsubseteq^{\text{prv}}$ are all reflexive and transitive (i.e., preorders), but not anti-symmetric (i.e., not partial orders). This is due to the fact that there exist channels that have "syntactic" differences but the same semantics; e.g., two channels having their columns swapped. However, if we are only interested in a specific type of leakage, then all channels such that $A \sqsubseteq B \sqsubseteq A$ (where $\sqsubseteq$ is one of $\sqsubseteq^{\text{avg}}, \sqsubseteq^{\text{max}}, \sqsubseteq^{\text{prv}}$) have identical leakage, so we can view them as the "same channel" (either by working on the equivalence classes of $\sqsubseteq \cup \sqsupseteq$ or by writing all channels in some canonical form).

Seeing now $\sqsubseteq$ as a partial order, the natural question is whether it forms a lattice that is whether suprema and infima exist. If it exists, the supremum $A \vee B$ has an interesting property: it is the "least safe" channel that is safer than both A and B (any channel C such that $A \sqsubseteq C$ and $B \sqsubseteq C$ would necessarily satisfy $A \vee B \sqsubseteq C$). If we wanted a channel that is safer than both A and B, $A \vee B$ would be a natural choice.

In this section, we briefly discuss this problem and show that—in contrast to $\sqsubseteq^{\mathrm{avg}}$—both $\sqsubseteq^{\max}$ and $\sqsubseteq^{\mathrm{prv}}$ do have suprema and infima (i.e., they form a lattice).

7.1. Average-Case Refinement

In the case of $\sqsubseteq^{\mathrm{avg}}$, "equivalent" channels are those producing the exact same hypers. However, even if we identify such channels, it is known [7] that two channels A, B do not necessarily have a *least upper bound* wrt $\sqsubseteq^{\mathrm{avg}}$, hence $\sqsubseteq^{\mathrm{avg}}$ does not form a lattice.

7.2. Max-Case Refinement

In the case of $\sqsubseteq^{\max}$, "equivalent" channels are those producing the same posteriors (or more generally the same convex hull of posteriors). However, in contrast to $\sqsubseteq^{\mathrm{avg}}$, if we identify such channels that is if we represent a channel only by the convex hull of its posteriors, then $\sqsubseteq^{\max}$ becomes a lattice.

First, note that given a finite set of posteriors $P = \{\delta^y | y\}$, such that $\pi \in ch\,\{\delta^y\}_y$, i.e., such that $\pi = \sum_y a_y \delta^y$, it is easy to construct a channel C producing each posterior δ^y with output probability a_y. It suffices to take $C_{x,y} := \delta^y_x a_y / \pi_x$.

Thus, $A \vee^{\max} B$ can be simply constructed by taking the intersection of the convex hulls of the posteriors of A, B. This intersection is a convex polytope itself, so it has (finitely many) extreme points, so we can construct $A \vee^{\max} B$ as the channel having exactly those as posteriors. $A \wedge^{\max} B$, on the other hand, can be constructed as the channel having as posteriors the union of those of A and B.

Note that computing the intersection of polytopes is NP-hard in general [21], so $A \vee^{\max} B$ might be hard to construct. However, efficient special cases do exist [22]; we leave the study of the hardness of $\vee^{\max}$ as future work.

7.3. Privacy-Based Refinement

In the case $\sqsubseteq^{\mathrm{prv}}$, "equivalent" channels are those producing the same induced metric, i.e., $d_A = d_B$. Representing channels only by their induced metric, we can use the fact that $\mathbb{M}\mathcal{X}$ does form a lattice under $\geq$. We first show that any metric can be turned into a corresponding channel.

Theorem 17. *For any metric $d{:}\,\mathbb{M}\mathcal{X}$, we can construct a channel C^d such that $d_{C^d} = d$.*

Then, $A \vee^{\mathrm{prv}} B$ will be simply the channel whose metric is $d_A \vee d_B$, where $\vee$ is the supremum in the lattice of metrics, and similarly for $\wedge^{\mathrm{prv}}$.

Note that the infimum of two metrics d_1, d_2 is simply the max of the two (which is always a metric). The supremum, however, is more tricky, since the min of two metrics is not always a metric: the triangle inequality might be violated. Thus, we first need to take the min of d_1, d_2, then compute its "triangle closure", by finding the shortest path between all pairs of elements, for instance using the well-known Floyd–Warshall algorithm.

8. Conclusions

We have investigated various refinement orders for mechanisms for information protection, combining the max-case perspective typical of DP and its variants with the robustness of the QIF approach. We have provided structural characterizations of these preorders and methods to verify them efficiently. Then, we have considered various DP mechanisms, and investigated the relation between the ε-based measurement of privacy and our orders. We have shown that, while within the same family of mechanisms, a smaller ε implies the refinement order, this is almost never the case for mechanisms belonging to different families.

Author Contributions: Conceptualization, K.C., N.F. and C.P.; methodology, K.C., N.F. and C.P.; software, K.C. and N.F.; validation, K.C. and N.F.; formal analysis, K.C. and N.F.; investigation, K.C. and N.F.; resources, C.P.; data curation, K.C. and N.F.; writing—original draft preparation, K.C. and N.F.; writing—review and editing, K.C., N.F. and C.P.; visualization, K.C. and N.F.; supervision, K.C. and C.P.; project administration, K.C. and C.P.; funding acquisition, C.P. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, Grant agreement No. 835294, by the project ANR-16-CE25-0011 REPAS, and by the Equipe Associée LOGIS.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proofs of Results about the Max-Case Refinement

Theorem A1. *Let $\pi\colon \mathbb{D}\mathcal{X}$. If $A \sqsubseteq^{\max} B$, then the posteriors of B (under π) are convex-combinations of those of A, that is,*

$$supp\,[\pi, B] \quad \subseteq \quad ch\,supp\,[\pi, A] \tag{A1}$$

Moreover, if (A1) *holds and π is full support, then $A \sqsubseteq^{\max} B$.*

Proof. Note that seeing π as a row vector, πA and πB are the output distributions of A and B, respectively. Denote by α^y and β^z the posteriors of $[\pi, A]$ and $[\pi, B]$, respectively; we have as many posteriors as the elements in the support of the output distributions, that is, for each $y\colon supp\ \pi A, z\colon supp\ \pi B$. (A1) can be written as

$$\forall z\colon supp\ \pi B.\ \left(\beta^z = \textstyle\sum_y c^z_y \alpha^y \quad \text{where} \quad \sum_y c^z_y = 1\right) \tag{A2}$$

The proof consists of two parts: first, we show that (A2) for uniform π is *equivalent* to $A \sqsubseteq^{\max} B$. Second, we show that (A2) for full-support π *implies* (A2) for any other prior.

For the first part, letting π be uniform, we show that (A2) is equivalant to $R\tilde{A} = \tilde{B}$. This is easy to see since since the y-th row of $\tilde{A}$ is α^y and the z-th row of $\tilde{B}$ is β^z. Hence, we can construct R from the convex coefficients, and vice versa, as $R_{z,y} = c^z_y$.

For the second part, let $\pi : \mathbb{D}\mathcal{X}$ be full-support and $\hat{\pi}\colon \mathbb{D}\mathcal{X}$ be arbitrary. Since $supp\ \hat{\pi} \subseteq supp\ \pi$, we necessarily have $supp\ \hat{\pi}\mathrm{C} \subseteq supp\ \hat{\pi}\mathrm{C}$ for any channel C. Assume that (A1) holds for π and let c^z_y be the corresponding convex coefficients. Fixing an arbitrary $z\colon supp\ \hat{\pi}B$, define

$$\hat{c}^z_y \quad := \quad c^z_y\, \frac{(\hat{\pi}A)_y}{(\pi A)_y}\, \frac{(\pi B)_z}{(\hat{\pi}B)_z} \tag{A3}$$

We first show that

$$\begin{aligned}
& \textstyle\sum_y c^z_y \frac{(\hat{\pi}A)_y}{(\pi A)_y} && \\
= \quad & \textstyle\sum_x \frac{\hat{\pi}_x}{\pi_x} \sum_y c^z_y \frac{\pi_x A_{x,y}}{(\pi A)_y} && \text{Expand } \hat{\pi}A\text{, rearrangement} \\
= \quad & \textstyle\sum_x \frac{\hat{\pi}_x}{\pi_x} \sum_y c^z_y \alpha^y_x && \text{Def. of } \alpha^y \\
= \quad & \textstyle\sum_x \frac{\hat{\pi}_x}{\pi_x} \beta^z_x && \text{(A1)} \\
= \quad & \textstyle\sum_x \frac{\hat{\pi}_x}{\pi_x} \frac{\pi_x B_{x,z}}{(\pi B)_z} && \text{Def. of } \beta^z \\
= \quad & \frac{(\hat{\pi}B)_z}{(\pi B)z} && \text{rearrangement}
\end{aligned}$$

From this, it follows that $\sum_y \hat{c}^z_y = 1$.

Finally, denote by $\hat{\alpha}^y$ and $\hat{\beta}^z$ the posteriors of $[\hat{\pi}, A]$ and $[\hat{\pi}, B]$, respectively; we show that (A1) holds for $\hat{\pi}$. Fixing $x\colon \mathcal{X}$, we have that

$$\begin{array}{lll}
 & \sum_y \hat{c}^z_y \hat{\alpha}^y_x & \\
= & \sum_y c^z_y \dfrac{(\hat{\pi}A)_y}{(\pi A)_y} \dfrac{(\pi B)_z}{(\hat{\pi}B)_z} \dfrac{\hat{\pi}_x A_{x,y}}{(\hat{\pi}A)_y} & \text{Def. of } c^z_y \text{ and } \hat{\alpha}^y \\
= & \frac{(\pi B)z}{(\hat{\pi}B)_z} \frac{\hat{\pi}_x}{\pi_x} \sum_y c^z_y \dfrac{\pi_x A_{x,y}}{(\pi A)_y} & \text{Def. of } d^z_y\text{, rearrangement} \\
= & \frac{(\pi B)z}{(\hat{\pi}B)_z} \frac{\hat{\pi}_x}{\pi_x} \sum_y c^z_y \alpha^y_x & \text{Def. of } \alpha^y \\
= & \frac{(\pi B)z}{(\hat{\pi}B)_z} \frac{\hat{\pi}_x}{\pi_x} \beta^z_x & \text{(A1)} \\
= & \frac{(\pi B)z}{(\hat{\pi}B)_z} \frac{\hat{\pi}_x}{\pi_x} \dfrac{\pi_x B_{x,z}}{(\pi B)_z} & \text{Def. of } \beta^y \\
= & \dfrac{\hat{\pi}_x B_{x,z}}{(\hat{\pi}B)_z} & \text{Rearrangement} \\
= & \hat{\beta}^z_x & \text{Def. of } \hat{\beta}^z
\end{array}$$

□

Theorem A2. *The orders $\sqsubseteq^{\max}$ and $\sqsubseteq^{\max}_{\mathbb{Q}}$ coincide.*

Proof. Fix some arbitrary π and denote by α^y and β^z the posteriors of $[\pi, A]$ and $[\pi, B]$, respectively. Assuming $A \sqsubseteq^{\max} B$, from Theorem A1, we get that each β^z can be written as a convex combination $\sum_y c^z_y \alpha^y$. Hence,

$$\begin{array}{lll}
 & V^{\max}[\pi, B] & \\
= & \max_z V(\beta^z) & \text{Def. of } V^{\max} \\
= & \max_z V(\sum_y c^z_y \alpha^y) & \text{Theorem A1} \\
\leq & \max_z \max_y V(\alpha^y) & \text{quasi-convexity of } V \\
= & \max_y V(\alpha^y) & \\
= & V^{\max}[\pi, A] &
\end{array}$$

from which $A \sqsubseteq^{\max}_{\mathbb{Q}} B$ follows.

Now, assume that $A \not\sqsubseteq^{\max} B$, let $S = \mathit{ch\,supp}\,[\pi, A] \subseteq \mathbb{D}\mathcal{X}$ and define a vulnerability function $V\colon \mathbb{Q}\mathcal{X}$ that maps every prior $\sigma\colon \mathbb{D}\mathcal{X}$ to its Euclidean distance from S, that is,

$$V(\sigma) \;\; := \;\; \min_{\sigma'\colon S} \|\sigma - \sigma'\|_2 \tag{A4}$$

Since S is a convex set, it is well known that $V(\sigma)$ is convex on σ (hence also quasi-convex). Note that $V(\sigma) = 0$ for all $\sigma \in S$ and strictly positive anywhere else.

By definition of S, we have that $\alpha^y \in S$ and hence $V(\alpha^y) = 0$ for all posteriors of A, as a consequence $V^{\max}[\pi, A] = 0$. On the other hand, since $A \not\sqsubseteq^{\max} B$, from Theorem A1, we get that there exists some posterior of B such that $\delta^z \notin S$. As a consequence, $V^{\max}[\pi, B] \geq V(\delta^z) > 0 = V^{\max}[\pi, A]$, which implies that $A \not\sqsubseteq^{\max}_{\mathbb{Q}} B$. □

Theorem A3. $\sqsubseteq^{\mathrm{avg}}$ *is strictly stronger than* $\sqsubseteq^{\max}$*.*

Proof. The "stronger" part is essentially the data-processing inequality for max-case vulnerability [9] (Prop. 14). To show it directly, assume that $A \sqsubseteq^{\text{avg}} B$, that is, $AR = B$ for some channel R, and define a channel S from $\mathcal{Z}$ to $\mathcal{Y}$ as

$$S_{z,y} \quad := \quad R_{y,z} \frac{\sum_x A_{x,y}}{\sum_x B_{x,z}} \tag{A5}$$

It is easy to check that S is a valid channel, i.e., that $\sum_y S_{z,y} = 1$ for all y. Moreover, we have that

$$\begin{array}{lll} & (S\tilde{A})_{z,x} & \\ = & \sum_y R_{y,z} \frac{\sum_x A_{x,y}}{\sum_x B_{x,z}} \frac{A_{x,y}}{\sum_x A_{x,y}} & \text{Def. of } S, \tilde{A} \\ = & \frac{\sum_y A_{x,y} R_{y,z}}{\sum_x B_{x,z}} & \text{Algebra} \\ = & \frac{B_{x,z}}{\sum_x B_{x,z}} & AR = B \\ = & \tilde{B}_{z,x} & \text{Def. of } \tilde{B} \end{array}$$

hence $A \sqsubseteq^{\max} B$.

The "strictly" part has already been shown in the body of the paper: The two matrices A and B in (14) provide an example in which $A \sqsubseteq^{\max} B$, while $B \not\sqsubseteq^{\max} A$.
□

Appendix B. Proofs of the Results about the Privacy-Based Refinement

Proposition A1. *The orders $\sqsubseteq^{\text{prv}}_{\mathbb{M}}$ and $\sqsubseteq^{\text{prv}}$ coincide.*

Proof. Assuming $A \sqsubseteq^{\text{prv}}_{\mathbb{M}} B$, recall that a channel C satisfies d-privacy iff $\text{Priv}_d(C) \leq 1$. Note also that $\text{Priv}_{d_C}(C) = 1$. Setting $d = d_A$, we get $1 = \text{Priv}_{d_A}(A) \geq \text{Priv}_{d_A}(B)$, which implies that B satisfies d_A-privacy, hence $A \sqsubseteq^{\max} B$.

Assuming $A \sqsubseteq^{\text{prv}} B$, to show that $A \sqsubseteq^{\text{prv}}_{\mathbb{M}} B$, it is equivalent to show that A satisfies d-privacy only if B also satisfies it. Let $d \in \mathbb{M}\mathcal{X}$, if A satisfies d-privacy, then $d \geq d_A \geq d_B$; hence, B also satisfies d-privacy, concluding the proof. □

Theorem A4. *$\sqsubseteq^{\max}$ is strictly stronger than $\sqsubseteq^{\text{prv}}$, which is strictly stronger than $\sqsubseteq^{\text{prv}}_d$.*

Proof. The "stronger" part is a direct consequence of the fact that $\text{Priv}_d(C)$ can be expressed as max-case capacity for a suitable vulnerability measure $V_d\colon \mathbb{Q}\mathcal{X}$ (more concretely, a consequence of Theorems 2, A6 and 7). This is discussed in detail in Section 4.7.

For the "strictly" part, consider the following counter-example:

A	y_1	y_2
x_1	0.8	0.2
x_2	0.4	0.6

B	y_1	y_2
x_1	0.4	0.6
x_2	0.8	0.2

(A6)

The only difference between A and B is that the two rows have been swapped. Hence, $d_A = d_B$, which implies $A \sqsubseteq^{\text{prv}} B \sqsubseteq^{\text{prv}} A$. However, the posteriors of A, B (for uniform prior) are (written in columns):

$$
\begin{array}{|c|cc|}
\hline
A & y_1 & y_2 \\
\hline
x_1 & 2/3 & 1/4 \\
x_2 & 1/3 & 3/4 \\
\hline
\end{array}
\qquad
\begin{array}{|c|cc|}
\hline
B & y_1 & y_2 \\
\hline
x_1 & 1/3 & 3/4 \\
x_2 & 2/3 & 1/4 \\
\hline
\end{array}
\tag{A7}
$$

Since $(3/4, 1/4)$ cannot be written as a convex combination of $(2/3, 1/3)$ and $(1/4, 3/4)$, and similarly $(1/4, 3/4)$ cannot be written as a convex combination of $(1/3, 2/3)$ and $(3/4, 1/4)$, from Theorem A1, we conclude that $A \not\sqsubseteq^{\max} B \not\sqsubseteq^{\max} A$. □

Theorem A5. *Let $f : \mathcal{X} \to \mathcal{Y}$ be any query and A, B be two mechanisms on $\mathcal{Y}$. If $A \sqsubseteq^{\text{prv}} B$, then $A \circ f \sqsubseteq^{\text{prv}} B \circ f$.*

Proof. Define $d_{A\circ f}(x_1, x_2) = d_A(f(x_1), f(x_2))$, and similarly for $d_{B\circ f}$. Then, we have:

$$d_{A\circ f}(x_1, x_2) = d_A(f(x_1), f(x_2)) \geq d_B(f(x_1), f(x_2)) = d_{B\circ f} \tag{A8}$$

□

Theorem A6. *$\mathcal{ML}_d^{+,\max}$ is always achieved on a uniform prior π^u. Namely*

$$\max_{\pi} \mathcal{L}_d^{+,\max}(\pi, C) \;=\; \mathcal{L}_d^{+,\max}(\pi^u, C) \;=\; V_d^{\max}[\pi^u, C] \tag{A9}$$

Proof. Fix π, C, and let (a, δ^y) and (b, ρ^y) be the outer and inners of $[\pi, V_d]$ and $[\pi^u, V_d]$, respectively. Since $\delta_x^y = C_{x,y}\pi_x a_y^{-1}$ and $\rho_x^y = C_{x,y}|\mathcal{X}|^{-1}b_y^{-1}$, we have that:

$$V_d(\delta^y) \;=\; \max_{x,x'} d^{-1}(x, x')|\ln \frac{C_{x,y}\pi_x}{C_{x',y}\pi_{x'}}| \quad \text{and} \tag{A10}$$

$$V_d(\rho^y) \;=\; \max_{x,x'} d^{-1}(x, x')|\ln \frac{C_{x,y}}{C_{x',y}}| \tag{A11}$$

Moreover, it holds that:

$$
\begin{aligned}
& V_d(\delta^y) \\
= \;& \max_{x,x'} d^{-1}(x, x')|\ln \tfrac{C_{x,y}\pi_x}{C_{x',y}\pi_{x'}}| \\
\leq \;& \max_{x,x'} d^{-1}(x, x')(|\ln \tfrac{C_{x,y}}{C_{x',y}}| + |\ln \tfrac{\pi_x}{\pi_{x'}}|) && \text{triangle inequality} \\
\leq \;& \max_{x,x'} (d^{-1}(x, x')|\ln \tfrac{C_{x,y}}{C_{x',y}}|) + && \text{independent max} \\
& \max_{x,x'} (d^{-1}(x, x')|\ln \tfrac{\pi_x}{\pi_{x'}}|) \\
= \;& V_d(\rho^y) + V_d(\pi)
\end{aligned}
$$

Finally, we have that:

$$
\begin{aligned}
& \mathcal{L}_d^{+,\max}(\pi, C) \\
= \;& \max_y V_d(\delta^y) - V_d(\pi)
\end{aligned}
$$

$$\begin{aligned}
&\leq \max_y \left(V_d(\rho^y) + V_d(\pi)\right) - V_d(\pi) \\
&= \max_y V_d(\rho^y) \\
&= V_d^{\max}[\pi^u, C] \\
&= \mathcal{L}_d^{+,\max}(\pi^u, C) \qquad \text{since } V_d(\pi^u) = 0
\end{aligned}$$

which concludes the proof. □

Theorem A7. *[DP as max-case capacity] C satisfies $\varepsilon \cdot d$-privacy iff $\mathcal{ML}_d^{+,\max}(C) \leq \varepsilon$. In other words: $\mathcal{ML}_d^{+,\max}(C) = \mathrm{Priv}_d(C)$.*

Proof. Let ρ^y denote the inners of $[\pi^u, V_d]$. From Theorem A6, we have that

$$\mathcal{ML}_d^{+,\max}(C) \leq \varepsilon \quad \text{iff} \quad V_d(\rho^y) \leq \varepsilon \quad \text{for all } y$$

which, from the definition of V_d, holds iff $C_{x,y} \leq e^{\varepsilon \cdot d(x,x')} C_{x',y}$ for all x, x', y. □

Appendix C. Proofs of the Results on the Refinement Verification

Proposition A2 (Projection theorem, [23] (Prop. 1.1.9))**.** *Let $C \subset \mathbb{R}^n$ be closed and convex and let $z \in \mathbb{R}^n$. There exists a unique $z^* \in C$ that minimizes $\|z - x\|_2$ over $x \in C$, called the projection of z on C. Moreover, a vector z^* is the projection of z on C iff*

$$(z - z^*) \cdot (x - z^*) \leq 0 \qquad \text{for all } x \in C \tag{A12}$$

Theorem A8. *Let B^* be the projection of B on $A^\uparrow$.*

1. *If $B = B^*$, then $A \sqsubseteq^{\text{avg}} B$.*
2. *Otherwise, let $G = B - B^*$. The gain function $g(w, x) = G_{x,w}$ provides a counter-example to $A \sqsubseteq^{\text{avg}} B$, which is $V_g(\pi^u, A) < V_g(\pi^u, B)$, for uniform π^u.*

Proof. (1) is immediate from the definition of $\sqsubseteq^{\text{avg}}$. For (2), we first show that

$$B \cdot G > B^* \cdot G \geq X \cdot G \qquad \text{for all } X \in A^\uparrow \tag{A13}$$

(in other words that $X \cdot G = B^* \cdot G$ is a hyperplane with normal G, separating B from $A^\uparrow$). For the left-hand inequality, we have $B \cdot G - B^* \cdot G = G \cdot G = \|G\|_2^2 > 0$. Moreover, since $A^\uparrow$ is closed and convex, from the projection theorem (Proposition A2), we get that $(B - B^*) \cdot (X - B^*) \leq 0$ for all $X \in A^\uparrow$, from which $B^* \cdot G \geq X \cdot G$ directly follows.

The proof continues similarly to the one of [7] (Theorem 9). We write posterior vulnerability (for uniform prior) as a maximization over all remapping strategies S_A, S_B for A, B respectively, namely

$$V_g(\pi, A) = \tfrac{1}{|\mathcal{X}|} \max_{AS_A \in A^\uparrow} AS_A \cdot G \tag{A14}$$

$$V_g(\pi, B) = \tfrac{1}{|\mathcal{X}|} \max_{BS_B \in B^\uparrow} BS_B \cdot G \tag{A15}$$

Then, $V_g(\pi^u, A) < V_g(\pi^u, B)$ follows from (A13) and the fact that $B \in B^\uparrow$ (the identity is a remapping strategy). □

Appendix D. Proofs of Results about Refinement Comparison

We call a truncated geometric mechanism 'square' if is has the same input and output space (that is, the channel representation is a square matrix).

We first show that geometric mechanisms and truncated geometric mechanisms are equivalent to square mechanisms under $\sqsubseteq^{\text{avg}}$.

Lemma A1. *Let G^ε be a geometric mechanism. Then, the reduced (abstract) channel form of G^ε is the square channel $(\mathcal{X}, \mathcal{X}, TG^\varepsilon)$.*

Proof. First, note that the square channel is obtained from the (infinite) geometric by summing up all the 'extra' columns (i.e., those columns in $\mathcal{Y} \setminus \mathcal{X}$). Now, note that these 'extra' columns are scalar multiples of each other, since each column has the form

$$\begin{bmatrix} \frac{(1-\alpha)\cdot\alpha^k}{1+\alpha} \\ \\ \frac{(1-\alpha)\cdot\alpha^{k-1}}{1+\alpha} \\ \\ \frac{(1-\alpha)\cdot\alpha^{k-2}}{1+\alpha} \\ \\ \dots \end{bmatrix} \tag{A16}$$

for increasing values of k. Thus, the 'summing up' operation is a valid reduction operation, and so the infinite geometric is reducible to the corresponding square channel. □

Lemma A2. *Let TG^ε be a truncated geometric mechanism. Then, the reduced (abstract) channel form of TG^ε is the square channel $(\mathcal{X}, \mathcal{X}, TG^\varepsilon)$.*

Proof. First, note that the truncated geometric is obtained from the infinite geometric by summing up columns at the ends of the matrix. This is exactly the 'reduction' step noted above. We can continue, as above, to sum up 'extra' columns until we get a square matrix. □

Corollary A1. *Any $\sqsubseteq^{\text{avg}}$ refinement that holds for a square geometric mechanism $(\mathcal{X}, \mathcal{X}, G^\varepsilon)$ also holds for any truncated geometric mechanism or (the) geometric mechanism G^ε having domain $\mathcal{X}$.*

Note that we only define truncation as far as the square geometric matrix, since at this point the columns of the matrix are linearly independent and can no longer be truncated via matrix reduction operations. We now show that refinement holds for the square geometric mechanisms.

Lemma A3. *Let TG^ε be a square geometric mechanism. Then, decreasing ε produces a refinement of it. That is, $TG^\varepsilon \sqsubseteq^{\text{avg}} TG^{\varepsilon'}$ iff $\varepsilon \geq \varepsilon'$.*

Proof. The square geometric mechanism TG^ε has the following form:

$$\begin{array}{c|cccc} TG^\varepsilon & x_1 & x_2 & \ldots & x_n \\ \hline x_1 & \frac{1}{1+\alpha} & \frac{\alpha\cdot(1-\alpha)}{1+\alpha} & \ldots & \frac{\alpha^{n-1}}{1+\alpha} \\ x_2 & \frac{\alpha}{1+\alpha} & \frac{1-\alpha}{1+\alpha} & \ldots & \frac{\alpha^{n-2}}{1+\alpha} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ x_n & \frac{\alpha^{n-1}}{1+\alpha} & \frac{\alpha^{n-2}\cdot(1-\alpha)}{1+\alpha} & \ldots & \frac{1}{1+\alpha} \end{array} \tag{A17}$$

where $\alpha = e^{-\varepsilon}$, and similarly for $TG^{\varepsilon'}$ with $\alpha = e^{-\varepsilon'}$.

Now, this matrix is invertible and the inverse has the following form:

$$\begin{array}{c|ccccc} (TG^\varepsilon)^{-1} & x_1 & x_2 & x_3 & x_4 & \ldots \\ \hline x_1 & \frac{1}{1-\alpha} & \frac{-\alpha}{1-\alpha} & 0 & 0 & \ldots \\ x_2 & \frac{-\alpha}{(1-\alpha)^2} & \frac{1+\alpha^2}{(1-\alpha)^2} & \frac{-\alpha}{(1-\alpha)^2} & 0 & \ldots \\ x_3 & 0 & \frac{-\alpha}{(1-\alpha)^2} & \frac{1+\alpha^2}{(1-\alpha)^2} & \frac{-\alpha}{(1-\alpha)^2} & \ldots \\ x_4 & 0 & 0 & \frac{-\alpha}{(1-\alpha)^2} & \frac{1+\alpha^2}{(1-\alpha)^2} & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \end{array} \tag{A18}$$

Recalling that

$$TG^\varepsilon \sqsubseteq^{\text{avg}} TG^{\varepsilon'} \text{ iff } TG^{\varepsilon'} = TG^\varepsilon R \tag{A19}$$

for some channel R, we can construct a suitable R using $(TG^\varepsilon)^{-1}$, namely $R = (TG^\varepsilon)^{-1} \cdot TG^{\varepsilon'}$. It suffices to show that R is a valid channel.

It is clear that the rows of R sum to 1, since it is the product of matrices with rows summing to 1.

Multiplying out the matrix R yields:

$$\begin{array}{c|cccc} R & x_1 & x_2 & x_3 & \ldots \\ \hline x_1 & \frac{1-\alpha\beta}{(1-\alpha)(1+\beta)} & \frac{(\beta-\alpha)(1-\beta)}{(1-\alpha)(1+\beta)} & \frac{\beta(\beta-\alpha)(1-\beta)}{(1-\alpha)(1+\beta)} & \ldots \\ x_2 & \frac{(1-\alpha\beta)(\beta-\alpha)}{(1-\alpha)^2(1+\beta)} & \frac{(1-2\alpha\beta+\alpha^2)(1-\beta)}{(1-\alpha)^2(1+\beta)} & \frac{(1-\alpha\beta)(1-\beta)(\beta-\alpha)}{(1-\alpha)^2(1-\beta)} & \ldots \\ x_3 & \frac{\beta(1-\alpha\beta)(\beta-\alpha)}{(1-\alpha)^2(1-\beta)} & \frac{(1-\alpha\beta)(1-\beta)(\beta-\alpha)}{(1-\alpha)^2(1-\beta)} & \frac{(1-2\alpha\beta+\alpha^2)(1-\beta)}{(1-\alpha)^2(1+\beta)} & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \end{array} \tag{A20}$$

where $\alpha = e^{-\varepsilon}$ and $\beta = e^{-\varepsilon'}$. The only way that any of these matrix entries can be less than 0 is if $\alpha > \beta$, or $\varepsilon < \varepsilon'$. Thus, R is a valid channel precisely when $\varepsilon \geq \varepsilon'$ and so $TG^\varepsilon \sqsubseteq^{\text{avg}} TG^{\varepsilon'}$ as required. □

The following theorems now follow from the previous lemmas.

Theorem A9. *Let $G^\varepsilon, G^{\varepsilon'}$ be geometric mechanisms. Then, $G^\varepsilon \sqsubseteq^{\text{avg}} G^{\varepsilon'}$ iff $\varepsilon \geq \varepsilon'$. That is, decreasing ε produces a refinement of the mechanism.*

Proof. Using Lemma A1, we can express G^{ε} as a square channel and from Lemma A3 it follows that the refinement holds. □

Theorem A10. *Let $TG^{\varepsilon}, TG^{\varepsilon'}$ be truncated geometric mechanisms. Then, $TG^{\varepsilon} \sqsubseteq^{\text{avg}} TG^{\varepsilon'}$ iff $\varepsilon \geq \varepsilon'$. That is, decreasing ε produces a refinement of the mechanism.*

Proof. As above, using Lemmas A2 and A3. □

Theorem A11. *Let $OTG^{\varepsilon}, OTG^{\varepsilon'}$ be over-truncated geometric mechanisms. Then, $OTG^{\varepsilon} \not\sqsubseteq^{\text{avg}} OTG^{\varepsilon'}$ for any $\varepsilon \neq \varepsilon'$. That is, decreasing ε does **not** produce a refinement.*

Proof. Consider the following counter-example:

A^{ε}	x_1	x_2
x_1	4/5	1/5
x_2	1/5	4/5
x_3	1/20	19/20

$B^{\varepsilon'}$	x_1	x_2
x_1	2/3	1/3
x_2	1/3	2/3
x_3	1/6	5/6

(A21)

Channels A^{ε} and $B^{\varepsilon'}$ are over-truncated geometric mechanisms parametrized by $\varepsilon = 2\log 2$, $\varepsilon' = \log 2$, respectively. We expect $B^{\varepsilon'}$ to be safer than A^{ε}, that is, $V_G[\pi^u, B^{\varepsilon'}] < V_G[\pi^u, A^{\varepsilon}]$. However, under the uniform prior π^u, the gain function

G	w_1	w_2
x_1	1/5	0
x_2	0	1
x_3	4/5	0

(A22)

yields $V_G[\pi^u, A^{\varepsilon}] = 0.33$ and $V_G[\pi^u, B^{\varepsilon'}] = 0.36$, thus $B^{\varepsilon'}$ leaks more than A^{ε} for this adversary. (In fact, for this gain function, we have $V_G[\pi^u, A^{\varepsilon}] = V_G(\pi^u)$ and so the adversary learns nothing from observing the output of A^{ε}). □

Theorem A12. *Let OTG^{ε} be an over-truncated geometric mechanism. Then, reducing ε does **not** produce a $\sqsubseteq^{\max}$ refinement; however, it **does** produce a $\sqsubseteq^{\text{prv}}$ refinement.*

Proof. We show the first part using a counter-example. Consider the following channels:

A^{ε}	x_1	x_2
x_1	4/5	1/5
x_2	1/5	4/5
x_3	1/20	19/20

$B^{\varepsilon'}$	x_1	x_2
x_1	2/3	1/3
x_2	1/3	2/3
x_3	1/6	5/6

(A23)

Channels A^{ε} and $B^{\varepsilon'}$ are over-truncated geometric mechanisms using $\varepsilon = 2\log 2$, $\varepsilon' = \log 2$, respectively. We can define the (prior) vulnerability V as the usual (convex) g-vulnerability. Then, under a uniform prior π^u, the gain function given by:

$$g(w, x_1) = \frac{1}{5} \tag{A24}$$

$$g(w, x_2) = -1 \tag{A25}$$

$$g(w, x_3) = \frac{4}{5} \tag{A26}$$

yields $V^{\max}[\pi^u, A] = 0$ and $V^{\max}[\pi^u, B] = \frac{2}{55}$. Thus, $A \not\sqsubseteq^{\max} B$.

For the second part, we first note that for any square geometric channel A^{ε} we have $d_A(x, x') = \varepsilon$ exactly when x, x' are adjacent rows in the matrix (this can be seen from the construction of the square channel). Now, the over-truncated geometric is obtained by summing columns of the square geometric. By construction, the square geometric A has adjacent elements $A_{x,y}, A_{x',y}$ satisfying ${}^{A_{x,y}}/{}_{A_{x',y}} = e^{\varepsilon}$ when $x > x'$ and x is *above* (or on) the diagonal of the channel matrix; otherwise, ${}^{A_{x,y}}/{}_{A_{x',y}} = e^{-\varepsilon}$. This means that each (over)-truncation step maintains the ${}^{A_{x,y}}/{}_{A_{x',y}}$ ratio except when x, y and x', y' occur on diagonal elements, in which case their sum is between $e^{-\varepsilon}$ and e^{ε}. Since this affects only two elements in each row, we still have that $d_A(x, x') = \varepsilon$ (until the final truncation step to produce a single 1 vector). Therefore, since $\sqsubseteq^{\text{prv}}$ holds for the square matrix, and it holds under truncation, we must have that it holds for over-truncated geometric mechanisms. Thus, reducing ε corresponds to refinement under $\sqsubseteq^{\text{prv}}$ as required. □

We show that the randomized response mechanism behaves well with respect to $\sqsubseteq^{\text{avg}}$ by considering three cases.

Firstly, we consider the case where $\mathcal{X} = \mathcal{Y}$. We use the following lemmas:

Lemma A4. *Let R^{α}, R^{β} be 'square' randomized response mechanisms. Then, $B = R^{\alpha}R^{\beta}$ is a randomized response mechanism with parameter $\varepsilon = log\frac{e^{\alpha+\beta}+k}{e^{\alpha}+e^{\beta}+k-1}$ where $k+1$ is the dimension of R^{α}, R^{β} and B.*

Proof. Observe that R^{α} can be factorised as $\frac{1}{e^{\alpha}+k}R$ where R has the form:

$$\begin{array}{c|cccc}
R & x_1 & x_2 & \dots & x_{k+1} \\
\hline
x_1 & e^{\alpha} & 1 & \dots & 1 \\
x_2 & 1 & e^{\alpha} & \dots & 1 \\
x_{k+1} & 1 & 1 & \dots & e^{\alpha}
\end{array} \tag{A27}$$

and similarly for R^{β}. Multiplying out gives the matrix:

$$\begin{array}{c|cccc}
B & x_1 & x_2 & \dots & x_{k+1} \\
\hline
x_1 & e^{\alpha+\beta} + k & e^{\alpha} + e^{\beta} + (k-1) & \dots & e^{\alpha} + e^{\beta} + (k-1) \\
x_2 & e^{\alpha} + e^{\beta} + (k-1) & e^{\alpha+\beta} + k & \dots & e^{\alpha} + e^{\beta} + (k-1) \\
x_{k+1} & e^{\alpha} + e^{\beta} + (k-1) & e^{\alpha} + e^{\beta} + (k-1) & \dots & e^{\alpha+\beta} + k
\end{array} \tag{A28}$$

(Note that the constant co-efficient factorised out the front does not affect the ε calculation for the channel). This is exactly the randomized response mechanism required. □

Lemma A5. *For any $a \geq 1, b \geq 0$, the function*

$$f(x) = \frac{ae^x + b}{e^x + a + b - 1} \tag{A29}$$

defined for $x \geq 0$ is increasing and has range $[1, a)$.

Proof. We can see that $f(x)$ is continuous for the given domain and the derivative $f'(x) = \frac{e^x(a-1)(a+b)}{(e^x+a+b-1)^2}$ is ≥ 0 for all $a \geq 1, b \geq 0$.

Additionally, at $x = 0$, the function is defined and equal to 1. In addition,

$$\lim_{x\to\infty} \frac{ae^x + b}{e^x + a + b - 1} = \lim_{x\to\infty} \frac{ae^x}{e^x} \tag{A30}$$

$$= a \tag{A31}$$

□

Lemma A6. *Let R^ε, $R^{\varepsilon'}$ be randomized response mechanisms represented by square matrices (that is, $\mathcal{X} = \mathcal{Y}$). Then, $R^\varepsilon \sqsubseteq^{\text{avg}} R^{\varepsilon'}$ iff $\varepsilon \geq \varepsilon'$.*

Proof. Note first that R^ε, $R^{\varepsilon'}$ are in reduced (abstract channel) form and so the partial order of $\sqsubseteq^{\text{avg}}$ holds.

From Lemma A4, we know that the composition of two randomized response mechanisms is another randomized response mechanism. Therefore, for the reverse direction, if $\varepsilon > \varepsilon'$ then Lemma A5 tells us that we can find a randomized response mechanism R' such that $R^\varepsilon R' = R^{\varepsilon'}$. In the case of equality, we can choose the identity mechanism.

For the forward direct, we show the contrapositive. If $\varepsilon < \varepsilon'$, then we know there exists an R such that $R^{\varepsilon'} R = R^\varepsilon$. However, this means that $R^{\varepsilon'} \sqsubseteq^{\text{avg}} R^\varepsilon$. Since the matrices are reduced (as channels), then $\sqsubseteq^{\text{avg}}$ is a partial order and so this implies $R^\varepsilon \not\sqsubseteq^{\text{avg}} R^{\varepsilon'}$. Thus, we must have $R^\varepsilon \sqsubseteq^{\text{avg}} R^{\varepsilon'} \implies \varepsilon \geq \varepsilon'$. □

Interestingly, we can use this result to show the second case where we consider 'over-truncated' mechanisms.

Lemma A7. *Let R^ε, $R^{\varepsilon'}$ be randomized response mechanisms with $\mathcal{Y} \subset \mathcal{X}$. Then, $R^\varepsilon \sqsubseteq^{\text{avg}} R^{\varepsilon'}$ iff $\varepsilon \geq \varepsilon'$.*

Proof. Notice that this 'over-truncated' randomized response mechanism is just a square randomized response mechanism 'glued' onto a matrix containing all values $\frac{1}{n+1}$.

Denote the corresponding square mechanisms by $S^\varepsilon, S^{\varepsilon'}$ (note the parameters are the same) and denote by N the matrix containing only $\frac{1}{n+1}$ (note that this is the same matrix for both R^ε and $R^{\varepsilon'}$).

From Lemma A6, we can find a square randomized response mechanism R satisfying $S^\varepsilon R = S^{\varepsilon'}$ iff $\varepsilon \geq \varepsilon'$. Notice that R must be doubly symmetric, since its ith row is the same as its ith column. Thus, the dot product of any column of R with any row vector containing only $\frac{1}{n+1}$ must yield $\frac{1}{n+1}$. In addition, we must have $N * R = N$. Now, we have that $R^\varepsilon R$ is just $S^\varepsilon R$ glued onto N which is the same as $S^{\varepsilon'}$ glued onto N. In addition, thus R also satisfies $R^\varepsilon R = R^{\varepsilon'}$. In addition, following the same arguments as for Lemma A6, we have $R^\varepsilon \sqsubseteq^{\text{avg}} R^{\varepsilon'}$ iff $\varepsilon \geq \varepsilon'$. □

Finally, we consider the case where $\mathcal{X} \subset \mathcal{Y}$.

Lemma A8. *Let R^ε, $R^{\varepsilon'}$ be randomized response mechanisms with $\mathcal{X} \subset \mathcal{Y}$. Then, $R^\varepsilon \sqsubseteq^{\text{avg}} R^{\varepsilon'}$ iff $\varepsilon \geq \varepsilon'$.*

Proof. The channel matrix R^ε is equivalent to the square randomized response mechanism S^ε of dimension $\|\mathcal{Y}\| \times \|\mathcal{Y}\|$ with the bottom $\|\mathcal{Y}\| - \|\mathcal{X}\|$ rows removed (and similarly for $R^{\varepsilon'}$). This means any solution R for S^ε is a solution for R^ε. Thus, for the reverse direction, if $\varepsilon \geq \varepsilon'$, we can always find an R such that $R^\varepsilon \sqsubseteq^{\text{avg}} R^{\varepsilon'}$.

For the forward direction, we prove the contrapositive. Note that in this case we cannot assume a partial order relation for $\sqsubseteq^{\text{avg}}$, since there may be columns of R^ε which are identical. If $\varepsilon < \varepsilon'$, we want to show that $R^\varepsilon \not\sqsubseteq^{\text{avg}} R^{\varepsilon'}$. To do this, we need to find a gain function and prior π such that $V_g(\pi, R^\varepsilon) < V_g(\pi, R^{\varepsilon'})$. The min-entropy leakage will do: this is simply the sum of the column maxima of the channel matrix. For the randomized response channels, this is given by

$$V(R^\varepsilon) = \frac{ae^\varepsilon + b}{e^\varepsilon + a + b - 1} \tag{A32}$$

for a channel with dimensions $a \times (a+b)$. This is an increasing function of ε (for $a \geq 1$), in fact the derivative is always positive for $a > 1$, hence we must have $\varepsilon < \varepsilon' \implies V(R^\varepsilon) < V(R^{\varepsilon'})$. Thus, $R^\varepsilon \not\sqsubseteq^{\text{avg}} R^{\varepsilon'}$. □

We can now conclude the following theorem from the main body of the paper:

Theorem A13. *Let R^ε be a randomized response mechanism. Then, decreasing ε in R produces a refinement. That is, $R^\varepsilon \sqsubseteq^{\text{avg}} R^{\varepsilon'}$ iff $\varepsilon \geq \varepsilon'$.*

Proof. Follows from Lemmas A6–A8. □

Theorem A14. *Let R be a randomized response mechanism, E an exponential mechanism and TG a truncated geometric mechanism. Then, $TG^\varepsilon \sqsubseteq^{\text{prv}} R^\varepsilon$ and $TG^\varepsilon \sqsubseteq^{\text{prv}} E^\varepsilon$. However, $\sqsubseteq^{\text{prv}}$ does not hold between E^ε and R^ε.*

Proof. We first show that $TG^\varepsilon \sqsubseteq^{\text{prv}} R^\varepsilon$, which is equivalent to showing that $d_R \leq d_{TG}$. For the geometric mechanism, we have $d_{TG}(x, x') = \varepsilon d(x, x')$ for all $x, x' \in \mathcal{X}$. For the randomized response mechanism, we have $d_R(x, x') = \varepsilon$ or $d_R(x, x') = 0$ (when $x, x' \notin \mathcal{Y}$). Thus, $d_R \leq d_{TG}$ and so $TG^\varepsilon \sqsubseteq^{\text{prv}} R^\varepsilon$.

We now show $TG^\varepsilon \sqsubseteq^{\text{prv}} E^\varepsilon$. Recall that we parametrize the exponential mechanism by the smallest possible ε such that it satisfies εd-privacy. In this case, we find that, for any pair x, x' we have $d_E(x, x') \leq \varepsilon d(x, x')$, whereas, for the geometric mechanism, we have $d_{TG}(x, x') = \varepsilon d(x, x')$. Therefore, $d_E \leq d_{TG}$ and so $TG^\varepsilon \sqsubseteq^{\text{prv}} E^\varepsilon$.

The proof of $\sqsubseteq^{\text{prv}}$ not holding between E^ε and R^ε was provided in the main body of the paper. □

Theorem A15. *Every (truncated geometric, randomized response, exponential) mechanism is 'the safest possible mechanism' when parametrized by $\varepsilon = 0$. That is, $L^\varepsilon \sqsubseteq^{\text{avg}} M^0$ for all mechanisms L, M (possibly from different families) and $\varepsilon > 0$.*

Proof. The intuition is that all channels parametrized by $\varepsilon = 0$ are equivalent to the **1** channel (that is, the $m \times 1$ channel consisting only of 1s). Indeed, the exponential and randomized response mechanisms parametrized by $\varepsilon = 0$ have every element equal to $\frac{1}{n}$ where $n = \|\mathcal{Y}\|$. These clearly reduce to the **1** channel. The truncated geometric mechanism contains all 0s except for the first and last column which contain $\frac{1}{2}$. Again, this reduces to the **1** channel. Since the **1** channel refines everything (that is, $L \sqsubseteq^{\text{avg}} \mathbf{1}$ for any channel L), the result follows. □

Appendix E. Proofs of the Results about the Lattice Properties

Theorem A16. *For any metric $d\colon \mathbb{M}\mathcal{X}$, we can construct a channel C^d such that $d_{C^d} = d$.*

Proof. Letting $d\colon \mathbb{M}\mathcal{X}$, we first show that, for any $x_0\colon \mathcal{X}$, we can construct a channel C^{x_0} whose induced metric is below d but coincides with it on all distances to x_0, that is:

$$d_{C^{x_0}} \leq d \qquad \text{and} \qquad d_{C^{x_0}}(x_0, x) = d(x_0, x) \;\; \text{for all } x\colon \mathcal{X} \tag{A33}$$

To construct C^{x_0}, we use just two outputs (i.e., $\mathcal{Y} = \{y_1, y_2\}$) and we use the fact that $\mathrm{tv}_{\otimes}$ on $\mathbb{D}\mathcal{Y}$ admits a *geodesic* that is a curve $\gamma : [0, +\infty] \to \mathbb{D}\mathcal{Y}$ such that

$$\mathrm{tv}_{\otimes}(\gamma(t), \gamma(t')) \quad = \quad |t - t'| \qquad \text{for all } t, t'\colon [0, +\infty] \tag{A34}$$

For instance, we can check that $\gamma(t) = (e^{-t-1}, 1 - e^{-t-1})$ is such a geodesic.

We can now use the geodesic, to assign probability distributions on each secret such that the properties (A33) are satisfied. Concretely, define each row x of C^{x_0} as:

$$C^{x_0}_{x,-} \quad := \quad \gamma(d(x_0, x)) \tag{A35}$$

We now check that the properties (A33) are satisfied:

$$\begin{array}{cll}
 & d_{C^{x_0}}(x_1, x_2) & \\
= & \mathrm{tv}_{\otimes}(\gamma(d(x_0, x_1)), \gamma(d(x_0, x_2))) & \text{Def. of } d_C, C^{x_0} \\
= & |d(x_0, x_1) - d(x_0, x_2)| & \gamma \text{ is a geodesic} \\
\leq & d(x_1, x_2) & \text{triangle ineq. for } d
\end{array}$$

and also

$$\begin{array}{cll}
 & d_{C^{x_0}}(x_0, x) & \\
= & \mathrm{tv}_{\otimes}(\gamma(d(x_0, x_0)), \gamma(d(x_0, x))) & \text{Def. of } d_C, C^{x_0} \\
= & |d(x_0, x_0) - d(x_0, x)| & \gamma \text{ is a geodesic} \\
= & d(x_0, x) & d(x_0, x_0) = 0
\end{array}$$

Finally, C^d is constructed as the visible choice of all $\{C^x\}_x$. As a consequence, d_{C^d} will be the max of the corresponding induced metrics $\{d_{C^x}\}_x$. From this and (A33), we can easily conclude that $d_{C^d} = d$.

Finally, note that the visible choice adds the columns of all mechanisms, so the constructed channel has $2|\mathcal{X}|$ columns. However, the equality of distances in (A33) is given by the *first column* of $C^{\tilde{x}}$ (this is because of the way γ is constructed), hence we can merge all second columns together, giving finally a simple construction for C^d with $\mathcal{Y} = \mathcal{X} \cup \{\perp\}$ (i.e., having $|\mathcal{X}| + 1$ columns)

$$C^d_{x,y} \quad = \quad |\mathcal{X}|^{-1} e^{-d(x,y)-1} \qquad\qquad x, y \in \mathcal{X} \tag{A36}$$

$$C^d_{x,\perp} \quad = \quad 1 - |\mathcal{X}|^{-1} \textstyle\sum_{y\colon \mathcal{X}} e^{-d(x,x')-1} \qquad\qquad x \in \mathcal{X} \tag{A37}$$

□

References

1. Dwork, C.; Mcsherry, F.; Nissim, K.; Smith, A. Calibrating noise to sensitivity in private data analysis. In Proceedings of the Third Theory of Cryptography Conference (TCC), New York, NY, USA, 4–7 March 2006; Lecture Notes in Computer Science; Halevi, S., Rabin, T., Eds.; Springer: Berlin/Heidelberg, Germany, 2006; Volume 3876, pp. 265–284.
2. Duchi, J.C.; Jordan, M.I.; Wainwright, M.J. Local Privacy and Statistical Minimax Rates. In Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science (FOCS), Berkeley, CA, USA, 26–29 October 2013; pp. 429–438. [CrossRef]
3. Chatzikokolakis, K.; Andrés, M.E.; Bordenabe, N.E.; Palamidessi, C. Broadening the scope of Differential Privacy using metrics. In Proceedings of the 13th International Symposium on Privacy Enhancing Technologies (PETS 2013), Bloomington, IN, USA, 10–12 July 2013; Lecture Notes in Computer Science; De Cristofaro, E., Wright, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; Volume 7981, pp. 82–102.
4. Köpf, B.; Basin, D.A. An information-theoretic model for adaptive side-channel attacks. In Proceedings of the 2007 ACM Conference on Computer and Communications Security (CCS 2007), Alexandria, VA, USA, 28–31 October 2007; Ning, P., di Vimercati, S.D.C., Syverson, P.F., Eds.; ACM: New York, NY, USA, 2007; pp. 286–296. [CrossRef]
5. Smith, G. On the Foundations of Quantitative Information Flow. In Proceedings of the 12th International Conference on Foundations of Software Science and Computation Structures (FOSSACS 2009), York, UK, 22–29 March 2009; de Alfaro, L., Ed.; Springer: York, UK, 2009; Volume 5504, pp. 288–302.
6. Alvim, M.S.; Chatzikokolakis, K.; Palamidessi, C.; Smith, G. Measuring Information Leakage Using Generalized Gain Functions. In Proceedings of the 25th IEEE Computer Security Foundations Symposium (CSF), Cambridge, MA, USA, 25–27 June 2012; pp. 265–279. [CrossRef]
7. McIver, A.; Morgan, C.; Smith, G.; Espinoza, B.; Meinicke, L. Abstract Channels and Their Robust Information-Leakage Ordering. In Proceedings of the Third International Conference on Principles of Security and Trust (POST), Grenoble, France, 5–13 April 2014; Lecture Notes in Computer Science; Abadi, M., Kremer, S., Eds.; Springer: Berlin/Heidelberg, Germany, 2014; Volume 8414, pp. 83–102.
8. Kifer, D.; Machanavajjhala, A. Pufferfish: A framework for mathematical privacy definitions. *ACM Trans. Database Syst.* **2014**, *39*, 3:1–3:36. [CrossRef]
9. Alvim, M.S.; Chatzikokolakis, K.; McIver, A.; Morgan, C.; Palamidessi, C.; Smith, G. Axioms for Information Leakage. In Proceedings of the 29th IEEE Computer Security Foundations Symposium (CSF), Lisbon, Portugal, 27 June–1 July 2016; pp. 77–92. [CrossRef]
10. Yasuoka, H.; Terauchi, T. Quantitative Information Flow—Verification Hardness and Possibilities. In Proceedings of the 23rd IEEE Computer Security Foundations Symposium, Edinburgh, UK, 17–19 July 2010; pp. 15–27. [CrossRef]
11. Malacaria, P. Algebraic foundations for quantitative information flow. *Math. Struct. Comput. Sci.* **2015**, *25*, 404–428. [CrossRef]
12. Clark, D.; Hunt, S.; Malacaria, P. Quantitative Information Flow, Relations and Polymorphic Types. *J. Log. Comput.* **2005**, *18*, 181–199. [CrossRef]
13. Malacaria, P. Assessing security threats of looping constructs. In Proceedings of the 34th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL 2007), Nice, France, 17–19 January 2007; Hofmann, M., Felleisen, M., Eds.; ACM: New York, NY, USA, 2007; pp. 225–235.
14. Landauer, J.; Redmond, T. A Lattice of Information. In Proceedings of theComputer Security Foundations Workshop VI, Franconia, NH, USA, 15–17 June 1993; pp. 65–70.
15. Hsu, J.; Gaboardi, M.; Haeberlen, A.; Khanna, S.; Narayan, A.; Pierce, B.C.; Roth, A. Differential Privacy: An Economic Method for Choosing Epsilon. In Proceedings of the IEEE 27th Computer Security Foundations Symposium, CSF 2014, Vienna, Austria, 19–22 July 2014; pp. 398–410. [CrossRef]
16. Ghosh, A.; Roth, A. Selling Privacy at Auction. In Proceedings of the 12th ACM Conference on Electronic Commerce, San Jose, CA, USA, 5–9 June 2011; ACM: New York, NY, USA, 2011; pp. 199–208. [CrossRef]
17. Andrés, M.E.; Bordenabe, N.E.; Chatzikokolakis, K.; Palamidessi, C. Geo-indistinguishability: differential privacy for location-based systems. In Proceedings of the 20th ACM Conference on Computer and Communications Security (CCS 2013), Berlin, Germany, 4–8 November 2013; ACM: New York, NY, USA, 2013; pp. 901–914. [CrossRef]

18. Barthe, G.; Köpf, B. Information-theoretic Bounds for Differentially Private Mechanisms. In Proceedings of the 24th IEEE Computer Security Foundations Symposium (CSF), Cernay-la-Ville, France, 27–29 June 2011; pp. 191–204.
19. Prasad Kasiviswanathan, S.; Smith, A. On the 'Semantics' of Differential Privacy: A Bayesian Formulation. *J. Priv. Confidentiality* **2008**, *6*. [CrossRef]
20. Chatzikokolakis, K.; Fernandes, N.; Palamidessi, C. Comparing systems: max-case refinement orders and application to differential privacy. In Proceedings of the 32nd IEEE Computer Security Foundations Symposium, Hoboken, NJ, USA, 25–28 June 2019; pp. 442–457. [CrossRef]
21. Tiwary, H.R. On the Hardness of Computing Intersection, Union and Minkowski Sum of Polytopes. *Discret. Comput. Geom.* **2008**, *40*, 469–479. [CrossRef]
22. Fukuda, K.; Liebling, T.M.; Lütolf, C. Extended Convex Hull. *Comput. Geom.* **2000**, *20*, 13–23. [CrossRef]
23. Bertsekas, D.P. *Convex Optimization Theory*; Athena Scientific: Belmont, MA, USA, 2009.

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.