Redrawing the Boundaries on Purchasing Data from Privacy-Sensitive Individuals

Kobbi Nissim
Ben-Gurion University and Harvard University.
Be'er Sheva, Israel and Cambridge, MA
kobbi@cs.bgu.ac.il

Salil Vadhan
Center for Research on Computation & Society and School of Engineering & Applied Sciences
Harvard University
Cambridge, MA
salil@seas.harvard.edu

David Xiao
CNRS
LIAFA and Université Paris 7
Paris, France
dxiao@liafa.univ-paris-diderot.fr

ABSTRACT

We prove new positive and negative results concerning the existence of truthful and individually rational mechanisms for purchasing private data from individuals with unbounded and sensitive privacy preferences. We strengthen the impossibility results of Ghosh and Roth (EC 2011) by extending it to a much wider class of privacy valuations. In particular, these include privacy valuations that are based on (ε, δ)-differentially private mechanisms for non-zero δ, ones where the privacy costs are measured in a per-database manner (rather than taking the worst case), and ones that do not depend on the payments made to players (which might not be observable to an adversary).

To bypass this impossibility result, we study a natural special setting where individuals have *monotonic privacy valuations*, which captures common contexts where certain values for private data are expected to lead to higher valuations for privacy (*e.g.* having a particular disease). We give new mechanisms that are individually rational for all players with monotonic privacy valuations, truthful for all players whose privacy valuations are not too large, and accurate if there are not too many players with too-large privacy valuations. We also prove matching lower bounds showing that in some respects our mechanism cannot be improved significantly.

Categories and Subject Descriptors

F.0 [**Theory of computation**]: General; K.4.1 [**Computers and society**]: Privacy

Keywords

differential privacy, mechanism design

ITCS'14, January 12–14, 2014, Princeton, New Jersey, USA.

ACM 978-1-4503-2698-8/14/01 ...$15.00.
http://dx.doi.org/10.1145/2554797.2554835.

1. INTRODUCTION

Computing over individuals' private data is extremely useful for various purposes, such as medical or demographic studies. Recent work on *differential privacy* first defined in [DMNS06, Dwo06] has focused on ensuring that analyses using private data can be carried out accurately while providing individuals a strong quantitative guarantee of privacy.

While differential privacy provides formal guarantees on how much information is leaked about an individual's data, it is silent about what incentivizes the individuals to share their data in the first place. A recent line of work [MT07, GR11, NST12, Xia13, NOS12, CCK+13, FL12, LR12, RS12] has begun exploring this question, by relating differential privacy to questions of mechanism design.

One way to incentivize individuals to consent to the usage of their private data is simply to pay them for using it. For example, a medical study may compensate its participants for the use of their medical data. However, determining the correct price is challenging: low payments may not draw enough participants, causing insufficient data for an accurate study, while high payments may be impossible for budgetary reasons.

Ghosh and Roth [GR11] approached this problem by allowing the mechanism to elicit *privacy valuations* from individuals. A privacy valuation is a description of how much disutility an individual experiences from having information about their private data revealed. By eliciting valuations, the mechanism is hopefully able to tailor payments to incentivize enough participants to produce an accurate result, while not paying too much.

1.1 The setting and previous work

We continue the study of purchasing private data from individuals as first proposed by Ghosh and Roth [GR11] (see [Rot12, PR13] for a survey of this area). Since we work in a game-theoretic framework, we will also call individuals "players". As in [GR11], we study the simple case where the private information consists of a single data bit, which players can refuse to provide but cannot modify (e.g. because the data is already certified in a trusted database, such as a medical record database).

To determine the price to pay players for their data bits, the mechanism elicits *privacy valuations* from them. We study the simple case where each player i's privacy valuation

is parameterized by a single real parameter v_i. For example, in Ghosh and Roth [GR11] they assume that player i loses $v_i\varepsilon$ utility when their data bit is used in an ε-differentially private mechanism. We will study a wider variety of privacy valuation functions in this paper. The valuations are known only to the players themselves, and therefore players may report false valuations if it increases their utility. Furthermore, because these valuations may be correlated with the data bits, the players may wish to keep their valuations private as well. It is instructive to keep in mind the application of paying for access to medical data (*e.g.* HIV status), where players cannot control the actual data bit, but their valuation might be strongly correlated to their data bit.

The goal of the mechanism is to approximate the sum of data bits while not paying too much. Based on the declared valuations, the mechanism computes payments to each of the players and obtains access to the purchased data bits from the players that accept the payments. The mechanism then computes and publishes an approximation to the sum of the data bits, which can cause the players some loss of privacy, which should be compensated for by the mechanism's payment.

The mechanism designer aims to achieve three goals, standard in the game theory literature: the mechanism should be individually rational, truthful, and accurate. A mechanism is *individually rational* if all players receive non-negative utility from participating in the game. In our context, this means that the mechanism is sufficiently compensating players for their loss in privacy, something that may be important for ethical reasons, beyond just incentivizing participation. Informally, a mechanism is *truthful* for player i on a tuple $x = (x_1, \ldots, x_n)$ of reports from the players if player i does not gain in utility by declaring some false type x'_i (while the other players' types remain unchanged). We aim to build mechanisms that are individually rational for all players, and truthful for as many players and inputs as possible (ideally for all players and inputs). A mechanism is *accurate* if the output of the mechanism is close to the true function it wishes to compute, in our case the sum of the data bits.

Ghosh and Roth [GR11] study the restricted setting (in their terminology the "insensitive value model") where players do not care about leaking their privacy valuations, as well as the general model (the "sensitive value model") where they may care and their valuations can be unbounded. They present two mechanisms in the insensitive value model, one that optimizes accuracy given a fixed budget and another that optimizes budget given a fixed accuracy constraint. They also prove that their mechanisms are individually rational and truthful under the assumption that each player i experiences a disutility of *exactly* $v_i\varepsilon$ when his data bit is used in an ε-differentially private mechanism.

In the general sensitive value model, they prove the following impossibility result: there is no individually rational mechanism with finite payments that can distinguish between the case where all players have data bit 0 and the case where all players have data bit 1.

This impossibility result spurred a line of work attempting to bypass it. Fleischer and Lyu [FL12] propose a Bayesian setting, where (for simplicity considering just Boolean inputs) there are publically known distributions D_0 and D_1 over privacy valuations, and each player who has data bit b_i receives a valuation v_i drawn from D_{b_i}. They show that in this model, it is possible to build a Bayes-Nash truthful, individually rational, and accurate mechanism.

In a related work, Roth and Schoenebeck [RS12] study a Bayesian setting where the agents' actual (dis)utilities are drawn from a known prior, and construct individually rational and ex-post truthful mechanism that are optimal for minimizing variance given a fixed budget and minimizing expected cost given a fixed variance goal. In comparison to [FL12], [RS12] studies a disutility value that does not quantitively relate to the privacy properties of the mechanism (but rather just a fixed, per-player disutility for participation), while it results in mechanisms satisfying a stronger notion of truthfulness.

Ligett and Roth [LR12] measure the privacy loss incurred from a player's decision to participate separately from the information leaked about the actual data (effectively ruling out arbitrary correlations between privacy valuations and data bits). They work in a worst-case (non-Bayesian) model and construct a mechanism that satisfies a relaxed "one-sided" notion of truthfulness and accuracy. However, their mechanism only satisfies individual rationality for players whose privacy valuation is not too high.

1.1.1 Improving the negative results

This line of work leaves several interesting questions open. The first is whether the impossibility result of [GR11] really closes the door on all meaningful mechanisms when players can have unbounded privacy valuations that can be arbitrarily correlated with their sensitive data.

There are two important loopholes that the result leaves open. First, their notion of privacy loss is pure ε-differential privacy, and they crucially use the fact that for pure ε-differentially private mechanisms the support of the output distribution must be identical for all inputs. This prevents their result from ruling out notions of privacy loss based on more relaxed notions of privacy, such as (ε, δ)-differential privacy for $\delta > 0$. As a number of examples in the differential privacy literature show, relaxing to (ε, δ)-differential privacy can be extremely powerful, even when δ is negligibly small but non-zero [DL09, HT10, DRV10, De12, BNS13]. Furthermore, even (ε, δ) differential privacy measures the worst-case privacy loss over all databases, and it may be the case that on most databases, the players' expected privacy loss is much less than the worst case bound.[1] Thus it is more realistic to use per-database measure of privacy loss (as done in [CCK+13]).

Second, the [GR11] notion of privacy includes as observable and hence potentially disclosive output the (sum of the) payments made to *all* the players, not just the sum of the data bits. This leaves open the possibility of constructing mechanisms for the setting where an outside observer is not able to to see some of the player's payments. For example, it may be natural to assume that, when trying to learn about

[1] For example, consider a mechanism that computes an ε-differentially private noisy sum of the first $n-1$ rows (which we assume are bits), and if the result is 0, also outputs a ε-differentially private noisy version of the n'th row (e.g. via "randomized response"). The worst case privacy loss for player n is ε. On databases of the form $(0, 0, \ldots, 0, b)$ the first computation results with 0 with probability $\approx \varepsilon$ and player n suffers ε privacy loss with this probability. However, if it is known that the database is very unlikely to be almost entirely zero, then player n may experience any privacy loss with only exponentially small probability.

player i, an observer learns the payments to all players *except* player i. In the extreme case, we could even restrict the outside observer to not see any of the payments, but only the approximation to the sum of the data bits. The Ghosh-Rosh impossibility proof fails in these cases. Indeed in this case where player i's own payment is not visible to the observer, there *does exist* an individually rational and accurate mechanism with finite payments: simply ask each player for their valuation v_i and pay them $v_i\varepsilon$, then output the sum of all the bits with noise of magnitude $O(1/\varepsilon)$. (The reason that this mechanism is unsatisfactory is that it is completely untruthful — players always gain by reporting a higher valuations.)

We will close both these gaps: our results will hold even under very mild conditions on how the players experience privacy loss (in particular capturing a per-database analogue of (ε, δ)-differential privacy), and even when *only* the approximate count of data bits is observable and *none* of the payments are observable.

1.1.2 Improving the positive results

Another question left open by the previous work is if we can achieve individual rationality and some form of truthfulness under a worst-case setting. Recall that [FL12] and [RS12] work in a Bayesian model, while [LR12] does not guarantee individual rationality for all players. Furthermore, in both [FL12] and [RS12] the priors are heavily used in *designing the mechanism*, and therefore their results break if the mechanism designer does not accurately know the priors. We will replace the Bayesian assumption with a simple qualitative assumption on the monotonicity of the correlation between players' data bits and their privacy valuation. For accuracy (but not individual rationality), we will assume a rough bound on how many players exceed a given threshold in their privacy valuations (similarly to [NOS12]).

Another question is the interpretation of the privacy loss functions. We observe that the truthfulness of the mechanisms in [GR11] crucially relies on the assumption that $v_i\varepsilon$ is the *exact* privacy loss incurred. As was argued by [NOS12] and [CCK+13], it seems hard to quantify the exact privacy loss a player experiences, as it may depend on the mechanism, all of the players' inputs, as well as an adversary's auxiliary information about the database. (See footnote 1 for an example.) It is much more reasonable to assume that the privacy valuations v_i declared by the players and the differential privacy parameter ε yield an *upper bound* on their privacy loss. When using this interpretation, the truthfulness of [GR11] no longer holds. The mechanisms we construct will remain truthful using the privacy loss function only as an upper bound on privacy loss (for players whose privacy valuations are not too large, similarly to the truthfulness guarantees of [NOS12, CCK+13, LR12]).

1.2 Our results

In our model there are n players labelled $1, \ldots, n$ each with a data bit $b_i \in \{0,1\}$ and a privacy valuation $v_i \in \mathbb{R}$, which we describe as a $2n$-tuple $(b, v) \in \{0,1\}^n \times \mathbb{R}^n$. The mechanism designer is interested in learning (an approximation of) $\sum b_i$. The players may lie about their valuation but they cannot lie about their data bit. A mechanism M is a pair of randomized functions $(M_{\mathsf{out}}, M_{\mathsf{pay}})$, where $M_{\mathsf{out}} : \{0,1\}^n \times \mathbb{R}^n \to \mathbb{Z}$ and $M_{\mathsf{pay}} : \{0,1\}^n \times \mathbb{R}^n \to \mathbb{R}^n$. Namely M_{out} produces an integer that should approximate $\sum b_i$ while M_{pay} produces payments to each of the n players.

Because the players are privacy-aware, the utility they derive from the game can be separated into two parts as follows:

$$\text{utility}_i = \text{payment}_i - \text{privacy loss}_i.$$

(Note that in this paper, we assume the players have no (dis)interest in the integer that M_{out} produces.) The privacy loss term will be quantified by a *privacy loss function* that depends on the identity of the player, his bit, his privacy valuation, and his declared valuation i, b, v, v'_i (where v'_i is not necessarily his true type v_i), the mechanism M, and the outcome (s, p) produced by $(M_{\mathsf{out}}, M_{\mathsf{pay}})$.

Strengthened impossibility result of non-trivial accuracy with privacy..

Our first result significantly strengthens the impossibility result of Ghosh-Roth [GR11].

THEOREM 1.1. *(Main impossibility result, informal. See Theorem 3.4.) Fix any mechanism M and any reasonable privacy loss functions. Then if M is truthful (even if only for players with privacy valuation 0) and individually rational and makes finite payments to the players (even if only when all players have privacy valuation 0), then M cannot distinguish between inputs $(b, v) = (0^n, 0^n)$ and $(b', v) = (1^n, 0^n)$.*

By "reasonable privacy loss functions," we mean that if from observing the output of the mechanism on an input (b, v), an adversary can distinguish the case that player i has data bit $b_i = 0$ from data bit $b_i = 1$ (while keeping all other inputs the same), then player i experiences a significant privacy loss (proportional to v_i) on database (b, v). In particular, we allow for a per-database notion of privacy loss. Moreover, we only need the adversary to be able to observe the mechanism's estimate of the count $\sum_j b_j$, and not any of the payments made to players. And our notion of indistinguishability captures not only pure ε-differential privacy but also (ε, δ)-differential privacy for $\delta > 0$. The conclusion of the result is as strong as conceivably possible, stating that M cannot distinguish between the two most different inputs (data bits all 0 vs. data bits all 1) even in the case where none of the players care about privacy.

We also remark that in our main impossibility result, in order to handle privacy loss functions that depend only on the distribution of the observable count and not the payment information, we crucially use the requirement that M be truthful for players with 0 privacy valuation. As we remarked earlier in Section 1.1.1 there exist M that are individually rational and accurate (but not truthful).

New notions of privacy and positive results..

One of the main conceptual contributions of this work is restricting our attention to a special class of privacy loss functions, which we use to bypass our main impossibility result. Essential to the definition of differential privacy (Definition 2.2) is the notion of *neighboring inputs*. Two inputs to the mechanism are considered neighboring if they differ only in the information of a single player, and in the usual notion of differential privacy, one player's information may differ arbitrarily. This view also characterized how previous work modeled privacy loss functions: in the sensitive value model of [GR11], the privacy loss function to a player i

on an input (b_i, v_i) was computed by considering how much changing to any possible neighbor (b_i', v_i') would affect the output of the mechanism. In contrast, we will restrict our attention to privacy loss functions that consider only how much changing to a specific subset of possible neighbors (b_i', v_i') would affect the output of the mechanism. By restricting to such privacy loss functions, we can bypass our impossibility results.

We now describe how we restrict (b_i', v_i'). Recall that in our setting a single player's type information is a pair (b_i, v_i) where $b_i \in \{0,1\}$ is a data bit and $v_i \in \mathbb{R}$ is a value for privacy. We observe that in many cases there is a natural sensitive value of the bit b_i, for example, if b_i represents HIV status, then we would expect that $b_i = 1$ is more sensitive than $b_i = 0$.

Therefore we consider only the following *monotonic valuations*: $(0, v_i)$ is a neighbor of $(1, v_i')$ iff $v_i \le v_i'$. Thus, if a player's true type is $(1, v_i')$, then he is only concerned with how much the output of the mechanism differs from the case that her actual type were $(0, v_i)$ for $v_i \le v_i'$.

Consider the pairs that we have excluded from consideration: any pairs $(b_i, v_i), (b_i, v_i')$ (*i.e.* the data bit does not change) and any pairs $(0, v_i), (1, v_i')$ where $v_i > v_i'$. By excluding these pairs we formally capture the idea that players are not concerned about revealing their privacy valuations *except* inasmuch as they may be correlated with their data bits b_i and therefore may reveal something about b_i. Since $b_i = 1$ is more sensitive than $b_i = 0$, the correlation says that privacy valuation when $b_i = 1$ should be larger than when $b_i = 0$. This can be seen as an intermediate notion between a model where players do not care at all about leaking their privacy valuation (the insensitive value model of [GR11]), and a model where players care about leaking any and all information about their privacy valuation (the sensitive value model of [GR11]).

Of course the assumption that players are not concerned about revealing their privacy valuation except inasmuch as it is correlated with their data is highly context-dependent. There may settings where the privacy valuation is intrinsically sensitive, independently of the players' data bits, and in these cases using our notion of monotonic valuations would be inappropriate. However, we believe that there are many settings where our relaxation is reasonable.

By using this relaxed notion of privacy, we are able to bypass our main impossibility result and prove the following:

THEOREM 1.2. *(Main positive result, informal. See Theorem 4.6.) For any fixed budget B and $\varepsilon > 0$, for privacy loss functions that only depend on how the output distribution changes between monotonic valuations, there exists a mechanism M that is individually rational for all players and truthful for players with low privacy valuation (specifically $v_i \le B/2\varepsilon n$). Furthermore, as long as the players with low privacy valuation do indeed behave truthfully, then regardless of the behavior of the players with high privacy valuation, the mechanism's output estimates the sum $\sum_i b_i$ to within $\pm(h + O(1/\varepsilon))$ where h is the number of players with high privacy valuation.*

Note that even though we fix a budget B beforehand and thus cannot make arbitrarily high payments, we still achieve individual rationality for all players, even those with extremely high privacy valuations v_i. We do so by ensuring that such players experience perfect privacy ($\varepsilon_i = 0$), assuming they have monotonic valuations. We also remark that while we do not achieve truthfulness for all players, this is not a significant problem as long as the number h of players with high privacy valuation is not too large. This is because the accuracy guarantee holds even if the non-truthful players lie about their valuations. We also give a small improvement to our mechanism that ensures truthfulness for all players with data bit 0, but at some additional practical inconvenience; we defer the details to the body of the paper.

We remark that besides our specific restriction to monotonic valuations in this paper, the underlying principle of studying restricted notions of privacy loss functions by considering only subsets of neighbors (where the subset should be chosen appropriately based on the specific context) could turn out to be a more generally meaningful and powerful technique that is useful to bypass impossibility results elsewhere in the study of privacy.

Lower bounds on accuracy..

The above positive result raises the question: can we adaptively select the budget B in order to achieve accuracy for all inputs, even those where some players have arbitrarily high privacy valuations? Recall that Theorem 1.1 does not preclude this because we are now only looking at monotonic valuations, whereas Theorem 1.1 considers arbitrary valuations. We nevertheless show that it is impossible:

THEOREM 1.3. *(Impossibility of accuracy for all privacy valuations, informal, see Theorem 5.3) For reasonable privacy loss functions that are only sensitive to changes in output distribution of monotonic neighbors, any M with finite payments that is truthful (even if only on players with 0 privacy valuation) and individually rational, there exist player privacy valuations v, v' such that M cannot distinguish between $(0^n, v)$ and $(1^n, v')$.*

The exact formal condition on finite payments is somewhat stronger here than in Theorem 1.1, but it remains reasonable; we defer the formal statement to the body of the paper.

Finally, we also prove a trade-off showing that when there is a limit on the maximum payment the mechanism makes, then accuracy cannot be improved beyond a certain point, even when considering only monotonic valuations. We defer the statement of this result to Section 5.2.

1.3 Related work

The relationship between differential privacy and mechanism design was first explored by [MT07]. Besides the already mentioned works, this relationship was explored and extended in a series of works [NST12], [Xia13] (see also [Xia11]), [NOS12], [CCK+13] (see also [CCK+11]), [HK12], [KPRU13]. In [MT07, NST12, HK12, KPRU13], truthfulness refers only to the utility that players derive from the outcome of the game (as in standard mechanism design) and differential privacy is treated as a separate property. The papers [Xia13, NOS12, CCK+13] study whether and when such mechanisms, which are separately truthful and differentially private, remain truthful even if the players are privacy-aware and may incur some loss in utility from the leakage of the private information. Differential privacy has also been used as a technical tool to solve problems that are not necessarily immediately obvious as being privacy-related; the original work of [MT07] does this, by using differential privacy to

construct approximately truthful and optimal mechanisms, while more recently, [KPRU13] use differential privacy as a tool to compute approximate equilibria. For more details, we refer the reader to the recent surveys of [Rot12, PR13].

Two ideas we draw on from this literature (particularly [NOS12, CCK+13]) are (1) the idea that privacy loss cannot be used as a threat because we do not know if a player will actually experience the maximal privacy loss possible, and therefore we should treat privacy loss functions only as upper bounds on the actual privacy loss, and (2) the idea that it is meaningful to construct mechanisms that are truthful for players with reasonable privacy valuations and accurate if most players satisfy this condition. Our mechanisms are truthful not for all players but only for players with low privacy valuation; it will be accurate if the mechanism designer knows enough about the population to set a budget such that most players have low privacy valuation (with respect to the budget).

2. DEFINITIONS

2.1 Notation

For two distributions X, Y we let $\Delta(X,Y)$ denote their total variation distance (i.e. statistical distance). For an integer i let $[i] = \{1,\ldots,i\}$. For any set S and any vector $v \in S^n$, we let $v_{-i} \in S^{n-1}$ denote the vector with entries $v_1,\ldots,v_{i-1},v_{i+1},\ldots,v_n$. We use the following convention: a vector of n entries consisting of $n-1$ variables or constants followed by an *indexed* variable denotes the vector of n entries with the last variable inserted at its index. For example *e.g.* $(0^{n-1},v_i)$ denotes the vector with all zeros except at the i'th entry, which contains v_i. Some notation about the setting regarding mechanisms etc. was already introduced in Section 1.2.

2.2 Differential privacy

DEFINITION 2.1. *Two inputs $(b,v),(b',v') \in \{0,1\}^n \times \mathbb{R}^n$ are* i-neighbors *if $b_j = b'_j$ and $v_j = v'_j$ for all $j \neq i$. They are* neighbors *if they are i-neighbors for some $i \in [n]$.*

DEFINITION 2.2. *A randomized function f is said to be* (ε,δ)-differentially private *if for all neighbors $(b,v),(b',v')$, it holds that for all subsets S of the range of f:*

$$\Pr[f(b,v) \in S] \leq e^{\varepsilon}\Pr[f(b',v') \in S] + \delta. \qquad (2.1)$$

We say f is ε-differentially private if it is $(\varepsilon,0)$-differentially private.

The symmetric geometric random variable $\mathrm{Geom}(\varepsilon)$ takes integer values with the probability mass function given by $\Pr_{x \leftarrow_R \mathrm{Geom}(\varepsilon)}[x=k] \propto e^{-\varepsilon|k|}$ for all $k \in \mathbb{Z}$. It is well-known and easy to verify that for $b \in \{0,1\}^n$, the output $\sum b_i + \mathrm{Geom}(\varepsilon)$ is ε-differentially private.

2.3 Privacy loss functions

A *privacy loss function* for player i is a real-valued function $\lambda_i^{(M)}(b,v,v'_i,s,p_{-i})$ taking as inputs the vectors of all player types b,v, player i's declaration v'_i (not necessarily equal to v_i), and a possible outcome $(s,p_{-i}) \in \mathbb{Z} \times \mathbb{R}^{n-1}$ of M. The function also depends on the mechanism $M = (M_{\mathsf{out}}, M_{\mathsf{pay}})$. Finally we define

$$\mathsf{Loss}_i^{(M)}(b,v,v'_i) = \mathbb{E}_{(s,p)\leftarrow_R M(b,(v_{-i},v'_i))}[\lambda_i^{(M)}(b,v,v'_i,s,p_{-i})]. \qquad (2.2)$$

Observe that we have excluded player i's own payment from the output, as we will assume that an outside observer cannot see player i's payment. We let M_{-i} denote the randomized function $M_{-i}(b,v) = (M_{\mathsf{out}}(b,v), M_{\mathsf{pay}}(b,v)_{-i})$.

We comment that, in contrast to [CCK+13], we allow $\lambda_i^{(M)}$ to depend on the player's declaration v'_i to model the possibility that a player's privacy loss depends on his declaration. Allowing this dependence only strengthens our positive results, while our negative results hold even if we exclude this dependence on v'_i. We remark that even if $\lambda_i^{(M)}$ doesn't depend on v'_i, then $\mathsf{Loss}_i^{(M)}$ will still depend on v'_i, since it is an expectation over the output distribution of $M_{\mathsf{out}}(b,(v_{-i},v'_i))$. (See Equation 2.2).)

Since the choice of a specific privacy loss function depends heavily on the context of the mechanism being studied, we avoid fixing a single privacy loss function and rather study several reasonable properties that privacy loss functions should have. Also, while we typically think of privacy valuation as being positive and privacy losses as positive, our definition does not exclude the possibility that players may *want* to lose their privacy, and therefore we allow privacy valuations (and losses) to be negative. Our impossibility results will only assume non-negative privacy loss, while our constructions handle possibly negative privacy loss functions as long as the *absolute value* of the privacy loss function is bounded appropriately.

2.4 Mechanism design criteria

DEFINITION 2.3. *A mechanism $M = (M_{\mathsf{out}}, M_{\mathsf{pay}})$ is said to be* $([\alpha,\alpha'],\beta)$-accurate *on an input $(b,v) \in \{0,1\}^n \times \mathbb{R}^n$ if, setting $\bar{b} = \frac{1}{n}\sum_{i=1}^n b_i$, it holds that*

$$\Pr[M_{\mathsf{out}}(b,v) \notin ((\bar{b}-\alpha)n, (\bar{b}+\alpha')n)] < \beta.$$

We say that M is (α,β)-accurate *on (b,v) if it is $([\alpha,\alpha],\beta)$-accurate.*

We define $\mathsf{Pay}_i^{(M)}(b,v) = \mathbb{E}_{p\leftarrow_R M_{\mathsf{pay}}(b,v)}[p_i]$.

DEFINITION 2.4. *Fix n, a mechanism M on n players, and privacy loss functions $\lambda_1^{(M)},\ldots,\lambda_n^{(M)}$. We say M is* individually rational *if for all inputs $(b,v) \in \{0,1\}^n \times \mathbb{R}^n$ and all $i \in [n]$:*

$$\mathsf{Pay}_i^{(M)}(b,v) \geq \mathsf{Loss}_i^{(M)}(b,v,v_i).$$

M is truthful for input (b,v) and player i *if for all v'_i it holds that*

$$\mathsf{Pay}_i^{(M)}(b,v) - \mathsf{Loss}_i^{(M)}(b,v,v_i) \geq \mathsf{Pay}^{(M)}(b,(v_{-i},v'_i)) - \mathsf{Loss}_i^{(M)}(b,v,v'_i)$$

M is simply truthful *if it is truthful for all inputs and all players.*

3. IMPOSSIBILITY OF NON-TRIVIAL ACCURACY WITH PRIVACY

We will use a notion of distinguishability that captures when a function leaks information about an input pertaining to a particular player.

DEFINITION 3.1. *An input $(b,v) \in \{0,1\}^n \times \mathbb{R}^n$ is said to be* δ-distinguishable *for player i with respect to a randomized function f if there is an i-neighbor (b',v') such that $\Delta(f(b,v), f(b',v')) \geq \delta$.*

We choose a notion based on statistical distance because it allows us to capture (ε, δ)-differential privacy even for $\delta > 0$. Namely, if there is an input $(b, v) \in \{0,1\}^n \times \mathbb{R}^n$ that is δ-distinguishable for player i with respect to f, then f cannot be (ε, δ')-differentially private for any ε, δ' satisfying $\delta > \delta' + e^{\varepsilon} - 1 \approx \delta' + \varepsilon$. However, note that, unlike differential privacy, δ-distinguishability is a *per-input* notion, measuring how much privacy loss a player can experience on a particular input (b, v), *not* taking the worst case over all inputs.

For our impossibility result we will require that any specified privacy loss should be attainable if the player's privacy valuation is large enough, as long as there is in fact a noticeable amount of information about the player's type being leaked (*i.e.* the player's input is somewhat distinguishable). Note that having unbounded privacy losses is necessary for having any kind of negative result. If the privacy losses were always upper-bounded by some value L, then a trivially truthful and individually rational mechanism would simply pay every player L and output the exact sum of data bits.

DEFINITION 3.2. *A privacy loss function $\lambda_i^{(M)}$ for a mechanism M and player i is* increasing for δ-distinguishability *if there exists a real-valued function T_i such that for all $\ell > 0$, $b \in \{0,1\}^n$ and $v_{-i} \in \mathbb{R}^{n-1}$, if $v_i \geq T_i(\ell, b, v_{-i})$ and if (b, v) is δ-distinguishable for player i with respect to M_{out}, then $\mathsf{Loss}_i^{(M)}(b, v, v_i) > \ell$.*

Notice that in our definition of increasing for δ-distinguishability we only consider distinguishability for M_{out} and not for $(M_{\mathsf{out}}, M_{\mathsf{pay}})$. Being able to handle this definition is what makes our impossibility rule out mechanisms even for privacy loss functions depending only on the distribution of M_{out}.

Definition 3.2 implies that the privacy loss functions are unbounded. We next define a natural property of loss functions, that for privacy-indifferent players privacy loss is not affected by the particular value reported for v_i.

DEFINITION 3.3. *A privacy loss function $\lambda_i^{(M)}$ for a mechanism M and player i* respects indifference *if whenever $v_i = 0$ it follows that $\mathsf{Loss}_i^{(M)}(b, v, v_i') = \mathsf{Loss}_i^{(M)}(b, v, v_i'')$ for all v_i', v_i''.*

THEOREM 3.4. *Fix a mechanism M, a number of players n, and non-negative privacy loss functions $\lambda_1^{(M)}, \ldots, \lambda_n^{(M)}$. Suppose that the $\lambda_i^{(M)}$ respect indifference, and are increasing for δ-distinguishability for some $\delta \leq \frac{1}{6n}$.*

Suppose that M that satisfies all of the following:

- *M is individually rational.*
- *M has finite payments when all players are privacy-indifferent, in the sense that for all $b \in \{0,1\}^n$ and all $i \in [n]$, it holds that $\mathsf{Pay}_i^{(M)}(b, 0^n)$ is finite.*
- *M is truthful for privacy-indifferent players, namely M is truthful for all inputs (b, v) and players i such that $v_i = 0$.*

Then it follows that M cannot have non-trivial accuracy in the sense that it cannot be $(1/2, 1/3)$-accurate on $(0^n, 0^n)$ and $(1^n, 0^n)$.

PROOF. Let $\mathsf{Pay}_i, \mathsf{Loss}_i, \lambda_i$ denote $\mathsf{Pay}_i^{(M)}$, $\mathsf{Loss}_i^{(M)}$, $\lambda_i^{(M)}$. By the assumption that M has finite payments when all players are privacy-indifferent, we can define

$$P = \max_{i \in [n], b \in \{0,1\}^n} \mathsf{Pay}_i(b, 0^n) < \infty.$$

By the assumption that all the λ_i are increasing for δ-indistinguishability, we may define a threshold

$$L = \max_{i \in [n], b \in \{0,1\}^n} T_i(P, b, 0^{n-1})$$

such that for all $i \in [n], b \in \{0,1\}^n, v_i \geq L$, if $(b, (0^{n-1}, v_i))$ is δ-distinguishable, then $\mathsf{Loss}_i(b, (0^{n-1}, v_i), v_i) > P$.

We construct a sequence of $2n+1$ inputs $x^{(1,0)}, x^{(1,1)}, x^{(2,0)}, x^{(2,1)}, \ldots, x^{(n,0)}, x^{(n,1)}, x^{(n+1,0)}$. In $x^{(1,0)}$, all players have data bit 0 and privacy valuation 0. That is, $x^{(1,0)} = (0^n, 0^n)$. From $x^{(i,0)}$, we construct $x^{(i,1)}$ by changing player i's data bit b_i from 0 to 1 and valuation v_i from 0 to L. From $x^{(i,1)}$, we construct $x^{(i+1,0)}$ by changing player i's valuation v_i back from L to 0 (but b_i remains 1). Thus,

$$\begin{aligned} x^{(i,0)} &= ((1^{i-1}, 0, 0^{n-i}), (0^{i-1}, 0, 0^{n-i})), \quad \text{and} \\ x^{(i,1)} &= ((1^{i-1}, 1, 0^{n-i}), (0^{i-1}, L, 0^{n-i})) \end{aligned}$$

In particular, $x^{(n+1,0)} = (1^n, 0^n)$. Define the hybrid distributions $H^{(i,j)} = M_{\mathsf{out}}(x^{(i,j)})$.

CLAIM 3.5. *For all $i \in [n]$, $\mathsf{Pay}_i(x^{(i,1)}) \leq \mathsf{Pay}_i(x^{(i+1,0)}) \leq P$.*

To prove this claim, we first note that all players have privacy valuation 0 in $x^{(i+1,0)}$, so $\mathsf{Pay}_i(x^{(i+1,0)}) \leq P$ by the definition of P. Since player i has privacy valuation 0 in $x^{(i+1,0)}$, we also know that privacy loss of player i in input $x^{(i+1,0)}$ is independent of her declaration (since λ_i respects indifference). If player i declares L as her valuation instead of 0, she would get payment $\mathsf{Pay}_i(x^{(i,1)})$. By truthfulness for privacy-indifferent players, we must have $\mathsf{Pay}_i(x^{(i,1)}) \leq \mathsf{Pay}_i(x^{(i+1,0)})$.

By the definition of L it follows that $x^{(i,1)}$ cannot be δ-distinguishable for player i with respect to M_{out}. Otherwise, this would contradict individual rationality because on input $x^{(i,1)}$ player i would have privacy loss $> P$ while only getting payoff $\leq P$.

Since $x^{(i,1)}$ is not δ-distinguishable for player i with respect to M_{out}, and because $x^{(i,1)}$ is an i-neighbor of $x^{(i,0)}$ as well as $x^{(i+1,0)}$, it follows that

$$\Delta(H^{(i,0)}, H^{(i,1)}) < \delta \text{ and } \Delta(H^{(i,1)}, H^{(i+1,0)}) < \delta \qquad (3.1)$$

Finally, since Equation 3.1 holds for all $i \in [n]$, and since $H^{(1,0)} = M_{\mathsf{out}}(0^n, 0^n)$ and $H^{(n+1,0)} = M_{\mathsf{out}}(1^n, 0^n)$, we have by the triangle inequality that

$$\Delta(M_{\mathsf{out}}(0^n, 0^n), M_{\mathsf{out}}(1^n, 0^n)) < 2n\delta$$

But since $\delta \leq 1/6n$, this contradicts the fact that M has non-trivial accuracy, since non-trivial accuracy implies that we can distinguish between the output of M_{out} on inputs $(0^n, 0^n)$ and $(1^n, 0^n)$ with advantage greater than $1/3$, simply by checking whether the output is greater than $n/2$. □

3.1 Using sub-sampling for privacy loss functions with low distinguishability

We comment that the $\delta \leq 1/6n$ bound in Theorem 3.4 is tight up to a constant factor. Indeed, if players do not incur

significant losses when their inputs are $O(1/n)$-distinguishable, then an extremely simple mechanisms based on subsampling can be used to achieve truthfulness, individual rationality, and accuracy with finite budget.

Namely, suppose that the privacy loss functions are such that if for all i, if player i's input is not C/n-distinguishable for some constant C, then regardless of v_i, the loss to player i is bounded by P. Then the following mechanism is truthful, individually rational, and accurate: pay all players P, select at random a subset A of size k for some $k < C$ from the population, and output $(n/|A|) \cdot \sum_{i \in A} b_i$. By a Chernoff Bound, this mechanism is $(\eta, 2e^{-\eta^2 k})$-accurate for all $\eta > 0$. By construction no player's input is C/n-distinguishable and therefore their privacy loss is at most P and the mechanism is individually rational. Finally mechanism is truthful since it behaves independently of the player declarations.

4. POSITIVE RESULTS

For our positive results, we will require the following natural property from our privacy loss functions. Recall that we allow the privacy loss functions $\lambda_i^{(M)}$ to depend on a player's report v_i', in addition the the player's true type. We require the dependence on v_i' to be well-behaved in that if changing declarations does not change the output distribution, then it also does not change the privacy loss.

DEFINITION 4.1. *A privacy loss function* $\lambda_i^{(M)}$ respects identical output distributions *if the following holds: for all* b, v, *if the distribution of* $M_{-i}(b, v)$ *is identical to the distribution of* $M_{-i}(b, (v_{-i}, v_i'))$, *then for all* s, p, *it holds that* $\lambda_i^{(M)}(b, v, v_i', s, p) = \lambda_i^{(M)}(b, v, v_i, s, p)$.

The above definition captures the idea that if what the privacy adversary can see (namely the output of M_{-i}) doesn't change, then player i's privacy loss should not change.

4.1 Monotonic valuations

We now define our main conceptual restriction of the privacy loss functions to consider only *monotonic valuations*.

DEFINITION 4.2. *Two player types* $(b_i, v_i), (b_i', v_i')$ *taking value in* $\{0,1\} \times \mathbb{R}$ *are said to be* monotonically related *iff* ($b_i = 0$, $b_i' = 1$, *and* $v_i \leq v_i'$) *or* ($b_i = 1$, $b_i' = 0$, *and* $v_i \geq v_i'$). *Two inputs* $(b, v), (b', v') \in \{0,1\}^n \times \mathbb{R}^n$ *are* monotonic i-neighbors *if they are i-neighbors and furthermore* $(b_i, v_i), (b_i', v_i')$ *are monotonically related. They are* monotonic neighbors *if they are monotonic i-neighbors for some* $i \in [n]$.

Following [CCK+13], we also make the assumption that the privacy loss functions on a given output (s, p_{-i}) are bounded by the amount of influence that player i's report has on the probability of the output:

DEFINITION 4.3. *A privacy loss function* $\lambda_i^{(M)}$ *is* bounded by differential privacy *if the following holds:*

$$\left|\lambda_i^{(M)}(b, v, v_i', s, p_{-i})\right| \leq v_i \cdot \left(\max_{(b_i'', v_i'')} \log \frac{\Pr[M_{-i}(b, v) = (s, p_{-i})]}{\Pr[M_{-i}((b_{-i}, b_i''), (v_{-i}, v_i'')) = (s, p_{-i})]} \right) \quad (4.1)$$

Input: $(b, v) \in \{0,1\}^n \times \mathbb{R}^n$. Auxiliary inputs: budget $B > 0$, privacy parameter $\varepsilon > 0$.

1. For all $i \in [n]$, set $b_i' = b_i$ if $2\varepsilon v_i \leq B/n$, otherwise set $b_i' = 0$.
2. Output $\sum_{i=1}^n b_i' + \text{Geom}(\varepsilon)$.
3. Pay B/n to player i if $2\varepsilon v_i \leq B/n$, else pay player i nothing.

ALGORITHM 4.5. *Mechanism for monotonic valuations*

A privacy loss function $\lambda_i^{(M)}$ *is* bounded by differential privacy for monotonic valuations *if Equation 4.1 holds but the maximum is restricted to be only over* (b_i'', v_i'') *monotonically related to* (b_i, v_i).

As noted and used in [CCK+13], the RHS in the above definition can be upper-bounded by the level of (pure) differential privacy, and the same holds for monotonic valuations:

FACT 4.4. *If* M_{-i} *is ε-differentially private for i-neighbors (*i.e. *Equation 2.1 holds for all i-neighbors) and* $\lambda_i^{(M)}$ *is bounded by differential privacy (even if only for monotonic valuations), then player i's privacy loss is bounded by* $v_i\varepsilon$ *regardless of other player types, player declarations, or outcomes.*

As hinted at in the definition of privacy loss functions bounded by differential privacy for monotonic valuations, one can define an analogue of differential privacy where we take the maximum over just monotonically related neighbors. However this notion is not that different from the original notion of differential privacy, since satisfying such a definition for some privacy parameter ε immediately implies satisfying (standard) differential privacy for privacy parameter 3ε, since every two pairs (b_i, v_i) and (b_i', v_i') are at distance at most 3 in the monotonic-neighbor graph. The monotonic neighbor notion becomes more interesting if we consider a further variant of differential privacy where the privacy guarantee ε_i afforded to an individual depends on her data (b_i, v_i) (e.g. $\varepsilon_i = 1/v_i$). We defer exploration of this notion to a future version of this paper.

4.2 Mechanism for monotonic valuations

The idea behind our mechanism for players with monotic valuations (Algorithm 4.5) is simply to treat the data bit as 0 (the insensitive value) for all players who value privacy too much.

THEOREM 4.6. *For privacy loss functions that are bounded by differential privacy for monotonic valuations and respect identical output distributions, the mechanism M in Algorithm 4.5 satisfies the following:*

1. *M is truthful for all players with $2\varepsilon v_i \leq B/n$.*
2. *M is individually rational for all players*
3. *Assume only that the truthful players described in Point 1 do indeed declare their true types. Letting η denote*

the fraction of players where $b_i = 1$ and $2\varepsilon v_i > B/n$, it holds that M is $([\eta+\gamma,\gamma], 2e^{-\varepsilon\gamma n})$-accurate.

PROOF. **Truthfulness for players with $2\varepsilon v_i \le B/n$:** if $2\varepsilon v_i \le B/n$, then declaring any $v_i' \le B/(2\varepsilon n)$ has no effect on the output of the mechanism, and so there is no change in utility since the privacy loss functions respect identical output distributions. If player i declares some $v_i' > B/(2\varepsilon n)$, then he loses B/n in payment. Because M_{-i} is ε-differentially private if for i-neighbors (recall we assume an observer cannot see the change in p_i) and we assumed that the privacy loss functions are bounded by differential privacy for monotonic valuations, it follows that player i's privacy loss has absolute value at most $2\varepsilon v_i$ under a report of v_i and under a report of v_i' (Fact 4.4). Thus, there is at most a change of $2\varepsilon v_i$ in privacy, which is not sufficient to overcome the payment loss of B/n.

Individual rationality: consider any vector of types b, v and any player i. If $v_i \le B/2\varepsilon n$ then player i receives payment B/n. By the hypothesis that the privacy loss functions are bounded by differential privacy for monotonic valuations, and because the mechanism is ε-differentially private, the privacy loss to player i is bounded by $\varepsilon v_i < B/n$ (Fact 4.4), satisfying individual rationality.

Now suppose that player i has valuation $v_i > \frac{B}{2\varepsilon n}$. In this case the payment is 0. The mechanism sets $b_i' = 0$, and for every (b_i'', v_i'') monotonically related to (b_i, v_i) the mechanism also sets $b_i' = 0$. Since the report of player i does not affect b_j' or the payment to player j for $j \ne i$, monotonic neighbors will produce the *exact* same output distribution of M_{-i}.

Therefore the privacy loss of player i is 0. Indeed, since the privacy loss function is bounded by differential privacy for monotonic valuations, we have:

$$\left|\lambda_i^{(M)}(b, v, v_i', s, p_{-i})\right| \le v_i \cdot \left(\max_{(b_i'', v_i'')} \log \frac{\Pr[M_{-i}(b, v) = (s, p_{-i})]}{\Pr[M_{-i}((b_{-i}, b_i''), (v_{-i}, v_i'')) = (s, p_{-i})]}\right) \tag{4.2}$$

$$= 0$$

where in Equation 4.2 the maximum is taken over (b_i'', v_i'') monotonically related to (b_i, v_i).

Accuracy: the bits of the $(1-\eta)$ fraction of truthful players and players with $b_i = 0$ are always counted correctly, while the bits of the η fraction of players with $b_i = 1$ and large privacy valuation $v_i \ge B/(2\varepsilon n)$ are either counted correctly (if they declare a value less than $B/(2\varepsilon n)$) or are counted as 0 (if they declare otherwise).

This means that $\overline{b'} = \sum_{i=1}^n b_i'$ and $\overline{b} = \sum_{i=1}^n b_i$ satisfy $\overline{b'} \in [\overline{b} - \eta n, \overline{b'}]$.

By the definition of symmetric geometric noise, it follows that (letting v' be the declared valuations of the players) it holds that

$$\Pr[|M_{\mathsf{out}}(b, v') - \overline{b'}| \ge \gamma n] < 2e^{-\varepsilon\gamma n}.$$

The theorem follows.

□

4.2.1 Achieving better truthfulness

We can improve the truthfulness of Theorem 4.6 to include all players with data bit 0.

THEOREM 4.7. *Let M' be the same as in Algorithm 4.5, except that* all *players with $b_i = 0$ are paid B/n, even those with large privacy valuations. Suppose that the $\lambda_i^{(M')}$ are bounded by differential privacy for monotonic valuations and also respect identical output distributions. Then the conclusions of Theorem 4.6 hold and in addition the mechanism is truthful for all players with data bit $b_i = 0$.*

Note that, unlike Algorithm 4.5, here the payment that the mechanism makes to players depends on their data bit, and not just on their reported valuation. This might make it impractical in some settings (e.g. if payment is needed before players give permission to view their data bits).

PROOF. Increasing the payments to the players with $b_i = 0$ and privacy valuation $v_i > \frac{B}{2\varepsilon n}$ does not hurt individual rationality or accuracy. We must however verify that we have not harmed truthfulness. Since players are not allowed to lie about their data bit, the same argument for truthfulness of players with $b_i = 1$ and $v_i \le B/(2\varepsilon n)$ remains valid. It is only necessary to verify that truthfulness holds for all players with $b_i = 0$.

Observe that for players with $b_i = 0$, the output distribution of the mechanism is identical regardless of their declaration for v_i. Therefore by the assumption that the $\lambda_i^{(M)}$ respect identical output distributions, changing their declaration does not change their privacy loss. Furthermore, by the definition of M' changing their declaration does not change their payment as all players with $b_i = 0$ are paid B/n. Therefore, there is no advantage to declaring a false valuation. □

We remark that the only setting where Theorem 4.7 is preferable to Theorem 4.6 is when knowing the true valuations is important beyond simply helping to achieve an accurate output; in particular, notice that M' as defined in Theorem 4.7 does not guarantee any better accuracy or any lower payments (indeed, it may make more payments than the original Algorithm 4.5).

5. LOWER BOUNDS

5.1 Impossibility of non-trivial accuracy for all privacy valuations with monotonic privacy

One natural question that Algorithm 4.5 raises is whether we can hope to adaptively set the budget B based on the valuations of the players and thereby achieve accuracy for all inputs, not just inputs where most players' privacy valuations are small relative to some predetermined budget. In this section we show that this is not possible, even when only considering players who care about privacy for monotonic neighbors.

DEFINITION 5.1. *An input $(b, v) \in \{0,1\}^n \times \mathbb{R}^n$ is δ-*monotonically distinguishable *for player i with respect to a randomized function f if there is a monotonic i-neighbor (b', v') such that $\Delta(f(b, v), f(b', v')) \ge \delta$.*

DEFINITION 5.2. *A privacy loss function $\lambda_i^{(M)}$ for a mechanism M and player i is* increasing for δ-monotonic distinguishability *if there exists a real-valued function T_i such that for all $\ell > 0$, $b \in \{0,1\}^n$ and $v_{-i} \in \mathbb{R}^{n-1}$, if $v_i \ge T_i(\ell, b, v_{-i})$ and if (b, v) is δ-monotonically distinguishable for player i with respect to M_{out}, then $\mathsf{Loss}_i^{(M)}(b, v, v_i) > \ell$.*

THEOREM 5.3. *Fix any mechanism M and any number of players n, and fix any non-negative privacy loss functions $\lambda_1^{(M)}, \ldots, \lambda_n^{(M)}$. Suppose that the $\lambda_i^{(M)}$ respect indifference and are increasing for δ-monotonic distinguishability for $\delta \leq \frac{1}{3n}$.*

Suppose M satisfies all the following:

- *M is individually rational.*
- *M always has finite payments, in the sense that for all $b \in \{0,1\}^n, v \in \mathbb{R}^n$ and all $i \in [n]$ it holds that $\mathsf{Pay}_i^{(M)}(b, v)$ is finite.*
- *M is truthful for privacy-indifferent players, as defined in Theorem 3.4*

Then M does not have non-trivial accuracy for all privacy valuations, i.e. *M cannot be $(1/2, 1/3)$-accurate on $(0^n, v)$ and $(1^n, v)$ for all $v \in \mathbb{R}^n$.*

PROOF. The argument follows the same outline as the proof of Theorem 3.4, *i.e.* by constructing a sequence of hybrid inputs and using truthfulness for privacy-indifferent players and individual rationality to argue that the neighboring hybrids must produce statistically close outputs. However, we have to take more care here because for the hybrids in this proof there is no uniform way to set the maximum payment P and threshold valuation L for achieving privacy loss $> P$ at the beginning of the argument, since here we allow the finite payment bound to depend on the valuations (whereas Theorem 3.4 only refers to the payment bound when all valuations are zero). Instead, we set P_i, L_i for the i'th hybrids in a way that depends on $L_{[i-1]} = (L_1, \ldots, L_{i-1})$.

As before, we define $2n+1$ hybrid inputs $x^{(1,0)}, x^{(1,1)}, x^{(2,0)}, x^{(2,1)}, \ldots, x^{(n,0)}, x^{(n,1)}, x^{(n+1,0)}$ inductively as follows. In $x^{(1,0)}$, all players have data bit 0 and privacy valuation 0. That is, $x^{(1,0)} = (0^n, 0^n)$. From $x^{(i,0)}$, we define $x^{(i,1)}$ by changing player i's data bit from 0 to 1. From $x^{(i,1)} = (b^{(i)}, v^{(i)})$, we define $P_i = \mathsf{Pay}_i(x^{(i,1)})$ to be the amount that player i is paid in $x^{(i,1)}$, and $L_i = T_i(P_i, b^{(i)}, v_{-i}^{(i)})$ to be a privacy valuation beyond which payment P_i does not compensate for δ-distinguishability (as promised by Definition 5.2). Then we define $x^{(i+1,0)}$ by increasing the valuation of player i from 0 to L_i. By induction, for $i = 1, \ldots, n+1$, we have

$$x^{(i,0)} = (1^{i-1}0^{n-i+1}, L_{[i-1]}0^{n-i+1}).$$

Define the distribution $H^{(i)} = M_{\mathsf{out}}(x^{(i,0)})$.

CLAIM 5.4. $\mathsf{Pay}_i(x^{(i+1,0)}) \leq \mathsf{Pay}_i(x^{(i,1)}) = P_i$

On input $x^{(i,1)}$, player i has privacy valuation 0, so his privacy loss is independent of his declaration (since λ_i respects indifference). Declaring L_i would change the input to $x^{(i+1,0)}$, so by truthfulness for privacy-indifferent players, we have $\mathsf{Pay}_i(x^{(i+1,0)}) \leq \mathsf{Pay}_i(x^{(i,1)})$.

By the definition of L_i, $x^{(i+1,0)}$ cannot be δ-monotonically distinguishable for player i with respect to M_{out}. Otherwise, this would contradict individual rationality because on input $x^{(i+1,0)}$ player i would have privacy loss greater than P_i while only getting a payoff of at most P_i (by Claim 5.4).

Since $x^{(i+1,0)}$ is not δ-monotonically distinguishable for player i with respect to M_{out}, and because $x^{(i,0)}$ is an i-monotonic neighbor of $x^{(i+1,0)}$, we therefore deduce that $\Delta(H^{(i-1)}, H^{(i)}) < \delta$. Finally, since this holds for all $i \in [n]$, the triangle inequality implies that $\Delta(H^{(0)}, H^{(n)}) < n\delta$. But since $\delta \leq 1/3n$, this implies that

$$\Delta(M_{\mathsf{out}}(0^n, 0^n), M_{\mathsf{out}}(1^n, L)) < 1/3$$

contradicting the fact that M has non-trivial accuracy for all privacy valuations. □

5.2 Tradeoff between payments and accuracy

One could also ask whether the accuracy of Theorem 4.6 can be improved, *i.e.* whether it is possible to beat $(\eta + \gamma, 2e^{-\varepsilon\gamma n})$-accuracy. We now present a result that, assuming the mechanism does not exceed a certain amount of payment, limits the best accuracy it can achieve. (We note however that this bound is loose and does not match our mechanism.)

In order to prove optimality we will require that the privacy loss functions be growing with statistical distance, a strictly stronger condition than being increasing for δ-distinguishability. However, a stronger requirement is unavoidable since one can invent contrived privacy loss functions that are increasing but for which one can achieve $(\eta, 0)$-accuracy by simply by outputting $\sum b_i'$ as constructed in Algorithm 4.5 without noise (while preserving the same truthfulness and individual rationality guarantees). Nevertheless, being growing with statistical distance for monotonic neighbors is compatible with being bounded by differential privacy for monotonic neighbors (*i.e.* there exist functions that satisfy both properties), and therefore the following result still implies limits to how much one can improve the accuracy of our positive result Theorem 4.6 for all privacy loss functions bounded by differential privacy for monotonic neighbors.

DEFINITION 5.5. *$\lambda_i^{(M)}(b, v, v_i', s, p_{-i})$ is* growing with statistical distance (for monotonic neighbors) *if:*

$$\mathsf{Loss}_i^{(M)}(b, v, v_i) \geq v_i \cdot \left(\max_{(b', v')} \Delta(M_{\mathsf{out}}(b, v), M_{\mathsf{out}}(b', v')) \right)$$

where the maximum is taken over (b', v') that are (monotonic) i-neighbors of (b, v).

THEOREM 5.6. *Fix a mechanism M, a number of players n, and privacy loss functions $\lambda_i^{(M)}$ for $i = 1, \ldots, n$. Suppose that the $\lambda_i^{(M)}$ respect indifference and are growing with statistical distance for monotonic neighbors.*

Suppose that M satisfies the following:

- *M is individually rational.*
- *There exists a maximum payment over all possible inputs that M makes to any player who declares 0 privacy valuation. Call this maximum value P.*
- *M is truthful for privacy-indifferent players as defined in Theorem 3.4.*

Then it holds that for any $\tau, \gamma, \eta > 0$ such that $\eta + 2\gamma \leq 1$, and any $\beta < \frac{1}{2} - \frac{P}{\tau}\gamma n$, the mechanism M cannot be $([\eta + \gamma, \gamma], \beta)$-accurate on all inputs where at most an η fraction of the players' valuations exceed τ.

PROOF. Fix any $\tau, \eta, \gamma > 0$ and any $\beta < \frac{1}{2} - \frac{P}{\tau}\gamma n$. We prove the theorem by showing that M cannot be $([\eta + \gamma, \gamma], \beta)$-accurate. Let $h = \eta n$ denote the number of players with high privacy valuation allowed.

Fix any $L \geq Ph/(1 - 2\frac{P}{\tau}\gamma n - 2\beta)$. Consider the following sequence of hybrid inputs. Let $x^{(1,0)} = (0^n, 0^n)$. From $x^{(i,0)}$,

define $x^{(i,1)}$ by flipping player i's data bit from 0 to 1. From $x^{(i,1)}$, define $x^{(i+1,0)}$ by increasing the valuation of player i from 0 to L if $i \in [h+1]$, or from 0 to τ if $i \in (h+1, h+2\gamma n+1]$. By induction, we have:

$$\forall i \in [h+1],$$
$$x^{(i,0)} = (1^{i-1}0^{n-i+1}, L^{i-1}0^{n-i+1})$$
$$\forall i \in (h+1, h+2\gamma n+1],$$
$$x^{(i,0)} = (1^{i-1}0^{n-i+1}, L^{h}\tau^{i-h-1}0^{n-i+1})$$

These are well-defined since $h+2\gamma = (\eta+2\gamma)n \leq n$. Define the hybrids $H^{(i,0)} = M_{\mathsf{out}}(x^{(i,0)})$. To analyze these hybrids, we use the following claims.

CLAIM 5.7. *For any input (b,v) where player i is paid at most P, it holds that (b,v) is not δ-distinguishable for monotonic neighbors for player i with respect to M_{out} for any $\delta \geq P/v_i$.*

Claim 5.7 holds because by individual rationality, it holds that the privacy loss does not exceed P. By the assumption that the privacy loss functions are growing with statistical distance for monotonic neighbors, it follows that $\Delta(M_{\mathsf{out}}(b,v), M_{\mathsf{out}}(b',v')) \leq P/v_i$ for all (b',v') monotonic neighbors of (b,v).

CLAIM 5.8. $\mathsf{Pay}_i(x^{(i+1,0)}) \leq \mathsf{Pay}_i(x^{(i,1)}) \leq P$.

As in the proof of Theorem 5.3, this claim holds because on input $x^{(i,1)}$, player $i+1$ has 0 privacy valuation, and so $\mathsf{Pay}_i(x^{(i,1)}) \leq P$ by our assumption that the mechanism pays at most P to players with 0 privacy valuation. The inequality $\mathsf{Pay}_i(x^{(i+1,0)}) \leq \mathsf{Pay}_i(x^{(i,1)})$ follows as in the proof of Theorem 5.3 from the truthfulness of the mechanism for privacy-indifferent players and by the fact that the privacy loss functions respect indifference.

We may apply Claim 5.7 to conclude that for all $i \in [h]$, since player i has valuation L in $x^{(i+1,0)}$, it therefore holds that $x^{(i+1,0)}$ cannot be (P/L)-distinguishable for monotonic neighbors for player i. Since $x^{(i,0)}, x^{(i+1,0)}$ are monotonic i-neighbors, it follows that $\Delta(H^{(i,0)}, H^{(i+1,0)}) < P/L$.

Repeating the same argument for all $i \in [h+1, h+2\gamma n]$ and using the fact that player i has valuation τ in $x^{(i+1,0)}$ for these i, it follows that $\Delta(H^{(i,0)}, H^{(i+1,0)}) < P/\tau$.

Combining the above using the triangle inequality and applying the definition of L, we deduce that

$$\Delta(H^{(1,0)}, H^{(h+2\gamma n+1,0)}) < \frac{\eta n P}{L} + \frac{2\gamma n P}{\tau} \leq 1-2\beta \quad (5.1)$$

For $i \in [n]$, define the open interval on the real line $A(i) = (i-1-(\eta+\gamma)n, i-1+\gamma n)$. Since the sum of the data bits in $x^{(i,0)}$ is exactly $i-1$, in order for M to be $([\eta+\gamma,\gamma],\beta)$-accurate, it is necessary that

$$\Pr[H^{(i)} \in A(i)] > 1-\beta \text{ for all } i \in [h+2\gamma n+1] \quad (5.2)$$

Observe that $A(1)$ and $A(h+2\gamma n+1)$ are disjoint. Therefore, Equation 5.1 implies that

$$\Pr[H^{(1,0)} \in A(1)] < \Pr[H^{(h+2\gamma n+1,0)} \in A(1)] + 1 - 2\beta$$

By Equation 5.2 it follows that $\Pr[H^{(h+2\gamma n+1,0)} \in A(1)] < \beta$ and therefore from the previous inequality we deduce that $\Pr[H^{(1,0)} \in A(1)] < 1-\beta$. But this contradicts Equation 5.2, and therefore it must be the case that M is not $([\eta+\gamma,\gamma],\beta)$-accurate. □

REMARK 5.9. *A different way to evaluate the accuracy guarantee of our mechanism, (the one taken in the work of Ghosh and Roth [GR11]) would be to compare it to the optimal accuracy achievable in the class of all envy-free mechanisms with budget B. However, in our context it is not clear how to define envy-freeness: while it is clear what it means for player i to receive player j's payment, it is not at all clear (without making further assumptions) how to define the privacy loss of player i as if he were treated like player j, since this loss may depend on the functional relationship between the player i's type and the output of the mechanism. Because of this, our mechanism may not envy-free (for reasonable privacy loss functions), and so we refrain from using envy-free mechanisms as a benchmark.*

6. ACKNOWLEDGEMENTS

K.N., S.V., and D.X., were supported in part by NSF grant CNS-1237235, a gift from Google, Inc., and a Simons Investigator grant to Salil Vadhan. K.N. was also supported by ISF grant (276/12). D.X. was also supported by the French ANR Blanc program under contract ANR-12-BS02-005 (RDAM project).

References

[BNS13] A. Beimel, K. Nissim, and U. Stemmer. Private Learning and Sanitization: Pure vs. Approximate Differential Privacy. In *Proceedings of the 17th International Workshop on Randomization and Computation (RANDOM '13)*, Lecture Notes in Computer Science, pages 363–378. Springer-Verlag, 21–23 August 2013.

[CCK+13] Y. Chen, S. Chong, I. A. Kash, T. Moran, and S. Vadhan. Truthful mechanisms for agents that value privacy. In *Proceedings of the fourteenth ACM conference on Electronic commerce*, EC '13, pages 215–232, New York, NY, USA, 2013. ACM.

[CCK+11] Y. Chen, S. Chong, I. A. Kash, T. Moran, and S. P. Vadhan. Truthful Mechanisms for Agents that Value Privacy. *CoRR*, abs/1111.5472, 2011.

[De12] A. De. Lower bounds in differential privacy. In *Theory of cryptography conference (TCC '12)*, volume 7194 of *Lecture Notes in Comput. Sci.*, pages 321–338. Springer, Heidelberg, 2012.

[Dwo06] C. Dwork. Differential privacy. In *In Proc. ICALP*, pages 1–12. Springer, 2006.

[DL09] C. Dwork and J. Lei. Differential privacy and robust statistics. In *STOC'09—Proceedings of the 2009 ACM International Symposium on Theory of Computing*, pages 371–380. ACM, New York, 2009.

[DMNS06] C. Dwork, F. Mcsherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *In Proc. of the 3rd TCC*, pages 265–284. Springer, 2006.

[DRV10] C. Dwork, G. Rothblum, and S. Vadhan. Boosting and Differential Privacy. In *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS '10)*, pages 51–60. IEEE, 23–26 October 2010.

[FL12] L. K. Fleischer and Y.-H. Lyu. Approximately optimal auctions for selling privacy when costs are correlated with data. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, EC '12, pages 568–585, New York, NY, USA, 2012. ACM.

[GR11] A. Ghosh and A. Roth. Selling privacy at auction. In *Proc. 12th EC*, EC '11, pages 199–208, New York, NY, USA, 2011. ACM.

[HT10] M. Hardt and K. Talwar. On the geometry of differential privacy. In *STOC'10—Proceedings of the 2010 ACM International Symposium on Theory of Computing*, pages 705–714. ACM, New York, 2010.

[HK12] Z. Huang and S. Kannan. The Exponential Mechanism for Social Welfare: Private, Truthful, and Nearly Optimal. In *Proc. FOCS '12*, pages 140–149, 2012.

[KPRU13] M. Kearns, M. M. Pai, A. Roth, and J. Ullman. Mechanism Design in Large Games: Incentives and Privacy. In *Proc. ITCS 2013*, 2013. Available at http://arxiv.org/abs/1207.4084.

[LR12] K. Ligett and A. Roth. Take it or Leave it: Running a Survey when Privacy Comes at a Cost. In *Proceedings of the 8th Workshop on Internet and Network Economics*, pages 378–391. Springer, 2012.

[MT07] F. McSherry and K. Talwar. Mechanism design via differential privacy. In *Proceedings of the 48th Annual Symposium on Foundations of Computer Science*. Citeseer, 2007.

[NOS12] K. Nissim, C. Orlandi, and R. Smorodinsky. Privacy-aware mechanism design. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, EC '12, pages 774–789, New York, NY, USA, 2012. ACM.

[NST12] K. Nissim, R. Smorodinsky, and M. Tennenholtz. Approximately optimal mechanism design via differential privacy. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, ITCS '12, pages 203–213, New York, NY, USA, 2012. ACM.

[PR13] M. M. Pai and A. Roth. Privacy and mechanism design. *SIGecom Exch.*, 12(1):8–29, June 2013.

[Rot12] A. Roth. Buying private data at auction: the sensitive surveyor's problem. *SIGecom Exch.*, 11(1):1–8, June 2012.

[RS12] A. Roth and G. Schoenebeck. Conducting truthful surveys, cheaply. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, EC '12, pages 826–843, New York, NY, USA, 2012. ACM.

[Xia11] D. Xiao. Is privacy compatible with truthfulness? Technical Report 2011/005, Cryptology ePrint Archive, 2011.

[Xia13] D. Xiao. Is privacy compatible with truthfulness? In *In Proc. ITCS 2013*, pages 67–86, 2013.