Take it or Leave it: Running a Survey when Privacy Comes at a Cost

Katrina Ligett[1] and Aaron Roth[2]

[1] Computing and Mathematical Sciences and Division of the Humanities and Social Sciences, Caltech.
[2] Department of Computer and Information Sciences, University of Pennsylvania.

Abstract. In this paper, we consider the problem of estimating a potentially sensitive (individually stigmatizing) statistic on a population. In our model, individuals are concerned about their privacy, and experience some *cost* as a function of their privacy loss. Nevertheless, they would be willing to participate in the survey if they were compensated for their privacy cost. These cost functions are not publicly known, however, nor do we make Bayesian assumptions about their form or distribution. Individuals are rational and will misreport their costs for privacy if doing so is in their best interest. Ghosh and Roth recently showed in this setting, when costs for privacy loss may be correlated with private types, if individuals value *differential privacy*, no individually rational direct revelation mechanism can compute any non-trivial estimate of the population statistic. In this paper, we circumvent this impossibility result by proposing a modified notion of how individuals experience cost as a function of their privacy loss, and by giving a mechanism which does not operate by direct revelation. Instead, our mechanism has the ability to randomly approach individuals from a population and offer them a take-it-or-leave-it offer. This is intended to model the abilities of a surveyor who may stand on a street corner and approach passers-by.

1 Introduction

Voluntarily provided data is a cornerstone of medical studies, opinion polls, human subjects research, and marketing studies. Suppose you are a researcher and you would like to collect data from a population and perform an analysis on it. Presumably, you would like your sample, or at least your analysis, to be representative of the underlying population. Unfortunately, individuals' decisions of whether to participate in your study may skew your data: perhaps people with an embarrassing medical condition are less likely to respond to a survey whose results might reveal their condition. Some data collectors, such as the US Census, can get around the issue of voluntary participation by legal mandate, but this is rare. How might we still get analyses that represent the underlying population?

Statisticians and econometricians have of course attempted to address selection and non-response bias issues. One approach is to assume that the effect of unobserved variables has mean zero. The Nobel-prize-winning Heckman correction method [1] and the related literature instead attempt to correct for non-random samples by formulating a theory for the probabilities of the unobserved variables and using the theorized distribution to extrapolate a corrected sample. The limitations of these approaches is precisely in the assumptions they make on the structure of the data. Is it possible to address these issues without needing to "correct" the observed sample, while simultaneously minimizing the cost of running the survey?

One could try to incentivize participation by offering a reward for participation, but this only serves to skew the survey in favor of those who value the reward over the costs of participating (e.g., hassle, time, detrimental effects of what the study might reveal), which again may not result in a representative sample. Ideally, you would like to be able to find out exactly how much you would have to pay each individual to participate in your survey (her "value", akin to a reservation price), and offer her exactly that much. Unfortunately, traditional mechanisms for eliciting player values truthfully are not a good match for this setting because a player's value may be correlated with her private information (for example, individuals with an embarrassing medical condition might want to be paid extra in order to reveal it). Standard mechanisms based on the revelation principle are therefore no longer truthful. In fact, Ghosh and Roth [2] showed that when participation costs can be arbitrarily correlated with private data, no direct revelation mechanism can simultaneously offer non-trivial accuracy guarantees and be individually rational for agents who value their privacy.

The present paper tackles this problem of conducting a survey on sensitive information when the costs of participating might be correlated with the information itself. In order to allow us to focus on the problem of incentivizing participation, we set aside the problem of *truthfulness*, and assume that once someone has decided to voluntarily participate in our survey, she must respond truthfully. This can most simply be justified by assuming that survey responses are verifiable or cannot easily be fabricated (e.g., the surveyor requires documentation of answers, or, more invasively, actually collects a blood sample from the participant). While the approach we present in this paper works well with such verifiable responses, in addition, our framework provides a formal "almost-truthfulness" guarantee, that the expected utility a participant could gain by lying in the survey is at most very small.

Motivated by the negative result of Ghosh and Roth [2], we move away from direct revelation mechanisms, to a framework where the surveyor is allowed to make "take-it-or-leave-it" offers to randomly sampled members of the underlying population. The simplest "take-it-or-leave-it" mechanism one might construct is simply to offer all sampled individuals the same low price in return for their participation in the survey (where participation might come with, e.g., a guarantee of differential privacy on their private data). If it turns out that this price is not high enough to induce sufficient rates of participation, one would double the price and restart the mechanism with a fresh sample of individuals, repeating until a target participation rate is reached (or the survey budget is exhausted).

The statistics released from the survey would then be based (perhaps in a differentially private manner) on the private information of the participants at the final (highest) price.

One might hope that such a simple doubling scheme would suffice to get "reasonable" participation rates at "reasonably" low cost. In order to deduce when take-it-or-leave-it offers will be accepted, though, we need a concrete model for how individuals value their privacy. Ghosh and Roth [2] provide such a model—essentially, they interpret the differential privacy parameter as the parameter governing individuals' costs. However, as we argue, this model can be problematic.

Our Results Our first contribution is to document the "paradox of differential privacy"—in Section 2, we observe that the manner in which Ghosh and Roth propose to model privacy costs results in clearly nonsensical behavioral predictions, even in a quite simple take-it-or-leave-it setting. In Section 5, we offer an alternative model for the value of privacy in multi-stage protocols, using the tools and language of differential privacy. We then, in Section 6, present a privacy-preserving variant of the simple "double your offer" algorithm above, and examine its ability to incentivize participation in data analyses when the subjects' value for their private information may be correlated with the sensitive information itself. We show that our simple mechanism allows us to compute accurate statistical estimates, addressing the survey problem described above, and we present an analysis of the costs of running the mechanism relative to a fixed-price benchmark.

2 The Paradox of Differential Privacy

Over the past decade, differential privacy has emerged as a compelling privacy definition, and has received considerable attention. While we provide formal definitions in Section 4, differential privacy essentially bounds the sensitivity of an algorithm's output to arbitrary changes in individual's data. In particular, it requires that the probability of *any* possible outcome of a computation be insensitive to the addition or removal of one person's data from the input. Among differential privacy's many strengths are (1) that differentially private computations are approximately truthful [3] (which gives the almost-truthfulness guarantee mentioned above), and (2) that differential privacy is a property of the *mechanism* and is independent of the input to the mechanism.

A natural approach taken by past work (e.g., [2]) in attempting to model the cost incurred by participants in a computation on their private data is to model individuals as experiencing cost as a function of the *differential privacy* parameter ε associated with the mechanism using their data. We argue here, however, that modeling an individual's cost for privacy loss solely as any function $f(\varepsilon)$ of the privacy parameter ε predicts unnatural agent behavior and incentives.

Consider an individual who is approached and offered a deal: she can participate in a survey in exchange for \$100, or she can decline to participate and walk away. She is given the guarantee that both her participation decision and her input to the survey (if she opts to participate) will be treated in an ε-differentially private manner. In the usual language of differential privacy, what does this mean? Formally, her input to the mechanism will be the tuple containing her participation decision and her private type. If she decides not to participate, the mechanism output is not allowed to depend on her private type, and switching her participation decision to "yes" cannot change the probability of any outcome by more than a small multiplicative factor. Similarly, fixing her participation decision as "yes", any change in her stated type can only change the probability of any outcome by a small multiplicative factor.

How should she respond to this offer? A natural conjecture is that she would experience a higher privacy cost for participating in the survey than not (after all, if she does not participate, her private type has *no* effect on the output of the mechanism – she need not even have provided it), and that she should weigh that privacy cost against the payment offered, and make her decision accordingly.

However, if her privacy cost is solely some function $f(\varepsilon)$ of the privacy parameter of the mechanism, she is actually incentivized to behave quite differently. Since the privacy parameter ε is *independent* of her input, her cost $f(\varepsilon)$ will be identical *whether she participates or not.* Indeed, her participation decision does not affect her privacy cost, and only affects whether she receives payment or not, and so she will always opt to participate in exchange for any positive payment.

We view this as problematic and as not modeling the true decision-making process of individuals: real people are unlikely to accept arbitrarily low offers for their private data. One potential route to addressing this "paradox" would be to move away from modeling the value of privacy solely in terms of an input-independent privacy guarantee. This is the approach taken by [4]. Instead, we retain the framework of differential privacy, but introduce a new model for how individuals reason about the cost of privacy loss. Roughly, we model individuals' costs as a function of the differential privacy parameter only of the portion of the mechanism they participate in, and assume they do not experience cost from the parts of the mechanism that process data that they have not provided (or that have no dependence on their data).

3 Related Work

In recent years, differential privacy [5] has emerged as the standard solution concept for privacy in the theoretical computer science literature. There is by now a very large literature on this fascinating topic, which we do not attempt to survey here, instead referring the interested reader to a survey by Dwork [6].

McSherry and Talwar [3] propose that differential privacy could itself be used as a *solution concept* in mechanism design (an approach later used by Gupta et al. [7] and others). They observe that any differentially private mechanism is approximately truthful, while simultaneously having some resilience to collusion. Using differential privacy as a solution concept (as opposed to dominant strategy truthfulness) they give improved results in a variety of auction settings.

This literature was extended by a series of elegant papers by Nissim, Smorodinsky, and Tennenholtz [8], Xiao [9], Nissim, Orlandi, and Smorodinsky [10], and Chen et al. [4]. This line of work observes ([8, 9]) that differential privacy does not lead to exactly truthful mechanisms, and indeed that manipulations might be easy to find, and then seeks to design mechanisms that are exactly truthful even when agents explicitly value privacy ([9, 10, 4]). Recently, Huang and Kannan show that the mechanism used by McSherry and Talwar is *maximal in range*, and so can be made exactly truthful through the use of payments [11].

Feigenbaum, Jaggard, and Schapira consider (using a different notion of privacy) how the implementation of an auction can affect how many bits of information are leaked about individuals' bids [12].

Most related to this paper is an orthogonal direction initiated by Ghosh and Roth [2], who consider the problem of a data analyst who wishes to buy data from a population for the purposes of computing an accurate estimate of some population statistic. Individuals experience cost as a function of their privacy loss (as measured by differential privacy), and must be incentivized by a truthful mechanism to report their true costs. In particular, [2] show that if individuals experience disutility as a function of differential privacy, and if costs for privacy can be arbitrarily correlated with private types, then no individually rational direct revelation mechanism can achieve any nontrivial accuracy. Fleischer and Lyu [13] overcome this impossibility result by moving to a Bayesian setting, in which costs are drawn from known prior distributions which depend on the individual's private data, and by proposing a relaxation of how individuals experience privacy cost. In this paper, we also overcome this impossibility result, but by an abandoning the direct revelation model in favor of a model in which a surveyor can approach random individuals from the population and offer them take-it-or-leave-it offers, and by introducing a slightly different model for how individuals experience cost as a function of privacy. In contrast to [13], our results allow for *worst-case* correlations between private data and costs for privacy, and do not require any Bayesian assumptions. Also in this line of work, Roth and Schoenebeck [14] consider the problem of deriving Bayesian optimal survey mechanisms for computing minimum variance unbiased estimators of a population statistic from individuals who have costs for participating in the survey. Although the motivation of this work is similar, the results are orthogonal. In the present paper, we take a prior-free approach and model costs for private access to data using the formalism of differential privacy. In contrast, [14] takes a Bayesian approach, assuming a known prior over agent costs, and does not attempt to provide any privacy guarantee, and instead only seeks to pay individuals for their participation.

4 Preliminaries

We model databases as an ordered multiset of elements from some universe X: $D \in X^*$ in which each element corresponds to the data of a different individidual. We call two databases *neighbors* if they differ in the data of only a single individual.

Definition 1. *Two databases of size n $D, D' \in X^n$ are* neighbors *with respect to individual i if for all $j \neq i \in [n]$, $D_j = D'_j$.*

We can now define *differential privacy.* Intuitively, differential privacy promises that the output of a mechanism does not depend too much on any single individual's data.

Definition 2 ([5]). *A randomized algorithm A which takes as input a database $D \in X^*$ and outputs an element of some arbitrary range R is ε_i-differentially private with respect to individual i if for all databases $D, D' \in X^*$ that are neighbors with respect to individual i, and for all subsets of the range $S \subseteq R$, we have:*

$$\mathbb{P}\mathrm{r}[A(D) \in S] \leq \exp(\varepsilon_i) Pr[A(D') \in S]$$

A is ε_i-minimally differentially private with respect to individual i if $\varepsilon_i = \inf(\varepsilon \geq 0)$ such that A is ε-differentially private with respect to individual i. When it is clear from context, we will simply write ε_i-differentially private to mean ε_i-minimally differentially private.

A simple and useful fact is that *post-processing* does not affect differential privacy guarantees.

Fact 41 *Let $A : X^* \to R$ be a randomized algorithm which is ε_i-differentially private with respect to individual i, and let $f : R \to T$ be an arbitrary (possibly randomized) function mapping the range of A to some abstract range T. Then the composition $g \circ f : X^* \to T$ is ε_i-differentially private with respect to individual i.*

A useful distribution is the *Laplace* distribution.

Definition 3 (The Laplace Distribution). *The Laplace Distribution with mean 0 and scale b is the distribution with probability density function: $Lap(x|b) = \frac{1}{2b}\exp(-\frac{|x|}{b})$. We will sometimes write $Lap(b)$ to denote the Laplace distribution with scale b, and will sometimes abuse notation and write $Lap(b)$ simply to denote a random variable $X \sim Lap(b)$.*

A fundamental result in data privacy is that perturbing low sensitivity queries with Laplace noise preserves ε-differential privacy.

Theorem 1 ([5]). *Suppose $f : X^* \to \mathbb{R}^k$ is a function such that for all adjacent databases D and D', $||f(D) - f(D')||_1 \leq 1$. Then the procedure which on input D releases $f(D) + (X_1, \ldots, X_k)$, where each X_i is an independent draw from a $Lap(1/\varepsilon)$ distribution, preserves ε-differential privacy.*

We consider a (possibly infinite) collection of individuals drawn from some distribution over types $\mathcal{D}$. There exists a finite collection of private types T. Each individual is described by a private type $t_i \in T$, as well as a nondecreasing cost function $c_i : \mathbb{R}_+ \to \mathbb{R}_+$ that measures her disutility $c_i(\varepsilon_i)$ for having her private type used in a computation with a guarantee of ε_i-differential privacy.

Agents interact with the mechanism as follows. The mechanism will be endowed with the ability to select an individual i uniformly at random (without replacement) from $\mathcal{D}$, by making a call to a *population oracle* $\mathcal{O}_{\mathcal{D}}$. Once an individual i has been sampled, the mechanism can present i with a *take-it-or-leave-it offer*, which is a tuple $(p_i, \varepsilon_i^1, \varepsilon_i^2) \in \mathbb{R}_+^3$. p_i represents an offered payment, and ε_i^1 and ε_i^2 represent two privacy parameters. The agent then makes her participation decision, which consists of one of two actions: she can *accept* the offer, or she can *reject* the offer. If she accepts the offer, she communicates her (verifiable) private type t_i to the auctioneer, who may use it in a computation which is ε_i^2-differentially private with respect to agent i. In exchange she receives payment p_i. If she rejects the offer, she need not communicate her type, and receives no payment. Moreover, the mechanism guarantees that the bit representing whether or not agent i accepts the offer is used only in an ε_i^1-differentially private way, regardless of her participation decision.

5 An Alternate Model of Privacy Costs

We model agents as caring only about the privacy of their private type t_i, but because of possible correlations between costs and types they may also experience a cost when information about their cost function $c_i(\varepsilon_i)$ is revealed. To capture this while avoiding Bayesian assumptions, we take the following approach.

Implicitly, there is a (possibly randomized) process f_i which maps a user's private type t to her cost function c_i, but we make no assumption about the form of this map. This takes a worst case view—i.e., we have no prior over individuals' cost functions. For a point of reference, in a Bayesian model, the function f_i would represent user i's marginal distribution over costs conditioned on her type.

When individual i is faced with a take-it-or-leave-it offer, her type may affect two computations: first, her participation decision (which may be a function of her type) is used in some computation A_1 which will be ε_i^1-differentially private. Then, if she accepts the offer, she allows her type to be used in some ε_i^2-differentially private computation, A_2.

We model individuals as *caring about the privacy of their cost function only insofar as it reveals information about their private type.* Because their cost function is determined as a function of their private type, if P is some predicate over cost functions, if $P(c_i) = P(f_i(t_i))$ is used in a way that guarantees ε_i-differential privacy, then the agent experiences a privacy loss of some $\varepsilon_i' \le \varepsilon_i$ (which corresponds to a disutility of some $c_i(\varepsilon_i') \le c_i(\varepsilon_i)$). We write $g_i(\varepsilon_i) = \varepsilon_i'$ to denote this correspondence between a given privacy level and the effective privacy loss due to use of the cost function at that level of privacy. For example, if f_i is a deterministic injective mapping, then $f_i(t_i)$ is as disclosive as t_i and so $g_i(\varepsilon_i) = \varepsilon_i$. On the other hand, if f_i produces a distribution independent of the user's type, then $g_i(\varepsilon_i) = 0$ for all ε_i. Note that by assumption, $0 \leq g_i(\varepsilon_i) \leq \varepsilon_i$ for all ε_i and g_i.

5.1 Cost Experienced from a Take-It-Or-Leave-It Mechanism

Definition 4. *A Private Take-It-Or-Leave-It Mechanism is composed of two algorithms, A_1 and A_2. A_1 makes offer $(p_i, \varepsilon_i^1, \varepsilon_i^2)$ to individual i and receives a binary* participation decision. *If player i* participates*, she receives a payment of p_i in exchange for her private type t_i. A_1 performs no computation on t_i. The privacy parameter ε_i^1 for A_1 is computed by viewing the input to A_1 to be the vector of participation decisions, and the output to be the number of individuals to whom offers were made, the offers $(p_i, \varepsilon_i^1, \varepsilon_i^2)$, and an ε_i^1-differentially private count of the number of players who chose to participate at the highest price we offer.*

Following the termination of A_1, a separate algorithm A_2 computes on the reported private types of these participating individuals and outputs a real number $\widehat{s}$. The privacy parameter ε_i^2 of A_2 is computed by viewing the input to be the private types of the participating agents, and the output as $\widehat{s}$.

We assume that agents have quasilinear utility (cost) functions: given a payment p_i, an agent i who declines a take-it-or-leave-it offer (and thus receives no payment) and whose participation decision is used in an ε_i^1-differentially private way experiences utility $u_i = -c_i(g_i(\varepsilon_i^1)) \geq -c_i(\varepsilon_i^1)$. An agent who accepts a take-it-or-leave-it offer and receives payment p, whose participation decision is used in an ε_i^1-differentially private way, and whose private type is subsequently used in an ε_i^2-differentially private way experiences utility $u_i = p_i - c_i(\varepsilon_i^2 + g_i(\varepsilon_i^1)) \geq p_i - c_i(\varepsilon_i^2 + \varepsilon_i^1)$, by a composition property of differential privacy.

Remark 1. This model captures the correct cost model in a number of situations. Suppose, for example, that costs have correlation 1 with types, and $c_i(\varepsilon) = \infty$ if and only if $t_i = 1$, otherwise $c_i(\varepsilon) \ll p_i$. Then, asking whether an agent wishes to accept an offer $(p_i, \varepsilon_i, \varepsilon_i)$ is equivalent to asking whether $t_i = 1$ or not, and those accepting the offer are in effect answering this question twice. In this case, we have $g_i(\varepsilon) = \varepsilon$. On the other hand, if types and costs are completely uncorrelated, then there is no privacy loss associated with responding to a take-it-or-leave-it offer. This is captured by setting $g_i(\varepsilon) = 0$.

Agents wish to maximize their utility, and so the following lemma is immediate:

Lemma 1. *A utility-maximizing agent i will accept a take-it-or-leave-it offer $(p_i, \varepsilon_i^1, \varepsilon_i^2)$ when $p_i \geq c_i(\varepsilon_i^1 + \varepsilon_i^2)$*

Proof. We simply compare the lower bound on an agent's utility when accepting an offer with an upper bound on an agent's utility when rejecting an offer to find that agent i will always accept when

$$p_i - c_i(\varepsilon_i^1 + \varepsilon_i^2) \geq 0.$$

Remark 2. Note that this lemma is tight exactly when agent types are uncorrelated with agent costs, i.e., when $g_i(\varepsilon) = 0$. When agent types are highly correlated with costs, then rejecting an offer becomes more costly, and agents may accept take-it-or-leave-it offers at lower prices.

We make no claims about how agents respond to offers $(p_i, \varepsilon_i^1, \varepsilon_i^2)$ for which $p_i < c_i(\varepsilon_i^1 + \varepsilon_i^2)$. Note that since agents can suffer negative utility even by rejecting offers, it is possible that they will accept offers that lead to experiencing negative utility. Thus, in our setting, take-it-or-leave-it offers do not necessarily result in participation decisions that *truthfully* reflect costs in the standard sense. Nevertheless, Lemma 1 will provide a strong enough guarantee for us of *one-sided truthfulness*: we can guarantee that rational agents will accept all offers that guarantee them non-negative utility.

Note that our mechanisms will satisfy only a relaxed notion of *individual rationality*: we have not endowed agents with the ability to avoid having been given a take-it-or-leave it offer, even if both options (taking or rejecting) would leave her with negative utility. Agents who reject take-it-or-leave-it offers can experience negative utility in our mechanism because their rejection decision is observed and used in a computation; we limit this negative utility and the corresponding deviation from individual rationality by treating their rejection decision in a differentially private manner. Once the take-it-or-leave-it offer has been presented, agents are free to behave selfishly. We feel that both of these relaxations (of truthfulness and individual rationality) are well motivated by real world mechanisms in which surveyors may approach individuals in public, and crucially, they are necessary in overcoming the impossibility result in [2].

Most of our analysis holds for arbitrary cost functions c_i, but we do a benchmark cost comparison assuming *linear* utility functions of the form $c_i(\varepsilon) = v_i\varepsilon$, for some quantity v_i.

5.2 Accuracy

Our mechanism is designed to be used by a data analyst who wishes to compute some statistic about the private type distribution of the population. Specifically, the analyst gives the mechanism some function $Q : T \to [0, 1]$, and wishes to compute $a = \mathbb{E}_{t_i \sim \mathcal{D}}[Q(t_i)]$, the average value that Q takes among the population of agents $\mathcal{D}$. The analyst wishes to obtain an *accurate* answer, defined as follows:

Definition 5. *A randomized algorithm, given as input access to a population oracle $\mathcal{O}_\mathcal{D}$ which outputs an estimate $M(\mathcal{O}_\mathcal{D}) = \widehat{a}$ of a statistic $a = \mathbb{E}_{t_i \sim \mathcal{D}}[Q(t_i)]$ is α-accurate if:*

$$\Pr[|\widehat{a} - a| > \alpha] < \frac{1}{3}$$

where the probability is taken over the internal randomness of the algorithm and the randomness of the population oracle.

The constant $\frac{1}{3}$ is arbitrary, and is fixed only for convenience. It can be replaced with any other constant value without qualitatively affecting our results.

5.3 Cost

We will evaluate the cost incurred by our mechanism using a bi-criteria benchmark: For a parametrization of our mechanism which gives accuracy α, we will compare our mechanism's cost to a benchmark algorithm that has perfect knowledge of each individual's cost function, but is constrained to make every individual the same take-it-or-leave-it offer (the same fixed price is offered to each person in exchange for some fixed ε'-differentially private computation on her private type) while obtaining $\alpha/32$ accuracy.[3] That is, the benchmark mechanism must be "envy-free", and may obtain better accuracy than we do, but only by a constant factor. On the other hand, the benchmark mechanism has several advantages: it has full knowledge of each player's cost, and need not be concerned about sample error. For simplicity, we will state our benchmark results in terms of individuals with linear cost functions.

6 Mechanism and Analysis

Due to space constraints, proofs can be found in the full version.

6.1 The Take-It-Or-Leave-It Mechanism

In this section we describe our mechanism. It is *not* a direct revelation mechanism, and instead is based on the ability to present take-it-or-leave-it offers to uniformly randomly selected individuals from some population. This is intended to model the scenario in which a surveyor is able to stand in a public location and ask questions or present offers to passers by (who are assumed to arrive randomly). Those passing the surveyor have the freedom to accept or reject the offer, but they cannot avoid having heard it.

Our mechanism consists of two algorithms. Algorithm 1, the Harassment Mechanism, is run on samples from the population with privacy guarantee ε_0, until it terminates at some final epoch $\widehat{j}$; and then Algorithm 2, the Estimation Mechanism, is run on $(\mathrm{AcceptedSet}_{\widehat{j}}, \mathrm{EpochSize}(\widehat{j}), \varepsilon_0)$. The Harassment Mechanism operates in epochs, wherin a large number of individuals are each offered the same price. The price we offer increases by a multiplicative $(1+\eta)$ in each epoch, for some fixed η . If a differentially private count of the number of players accepting the offer in a given epoch is high enough, we call this the final epoch, and hand the participating individuals over to the Estimation Mechanism. The Estimation Mechanism then computes a differentially private (noisy) version of the desired statistic over this set of individuals who chose to participate at the highest price.

6.2 Privacy

Note that our mechanism offers the same ε_0 in every take-it-or-leave-it offer.

[3] Note that we have made no attempt to optimize the constant.

Algorithm 1 Algorithm A_1, the "Harassment Mechanism". It is parametrized by an accuracy level α, and we view its *input* to be the participation decision of each individual approached with a take-it-or-leave-it offer, and its *observable output* to be the number of individuals approached, the payments offered, and the noisy count of the number of players who accepted the offer in the final epoch.

Let $\text{EpochSize}(j) \leftarrow \frac{100(\log j+1)}{\alpha^2}$.
Let $j \leftarrow 1$.
Let $\varepsilon_0 = \alpha$
while TRUE do
  Let $\text{AcceptedSet}_j \leftarrow \emptyset$ and $\text{NumberAccepted}_j \leftarrow 0$ and $\text{Epoch}_j \leftarrow \emptyset$
  for $i = 1$ to $\text{EpochSize}(j)$ **do**
    Sample a new individual x_i from $\mathcal{D}$.
    Let $\text{Epoch}_j \leftarrow \text{Epoch}_j \cup \{x_i\}$.
    Offer x_i the take-it-or-leave it offer $(p_j, \varepsilon_0, \varepsilon_0)$ with $p_j = (1+\eta)^j$
    if i accepts **then**
      Let $\text{AcceptedSet}_j \leftarrow \text{AcceptedSet}_j \cup \{x_i\}$ and
        $\text{NumberAccepted}_j \leftarrow \text{NumberAccepted}_j + 1$.
  Let $\nu_j \sim \text{Lap}(1/\varepsilon_0)$ and $\text{NoisyCount}_j = \text{NumberAccepted}_j + \nu_j$
  if $\text{NoisyCount}_j \geq (1-\alpha/8)\text{EpochSize}(j)$ **then**
    Call Estimate$(\text{AcceptedSet}_j, \text{EpochSize}(j), \varepsilon_0)$.
  else
    Let $j \leftarrow j+1$

Algorithm 2 The Estimation Mechanism. We view its *inputs* to be the private types of each participating individual from the final epoch, and its *output* is a single numeric estimate.

Estimate$(\text{AcceptedSet}, \text{EpochSize}, \varepsilon)$:
  Let $\widehat{a} = \sum_{x_i \in \text{AcceptedSet}} Q(x_i) + \text{Lap}(1/\varepsilon)$
  Output $\widehat{a}/\text{EpochSize}$.

Theorem 2. *The Harassment Mechanism is ε_0-differentially private with respect to the participation decision of each individual approached.*

Theorem 3. *The Estimation Mechanism is ε_0-differentially private with respect to the participation decision and private type of each individual approached.*

Note that these two theorems, together with Lemma 1, imply that each agent will accept her take-it-or-leave-it offer of $(p_j, \varepsilon_0, \varepsilon_0)$ whenever $p_j \geq c_i(2\varepsilon_0)$.

6.3 Accuracy

Theorem 4. *Our overall mechanism, which first runs the Harassment Mechanism and then hands the types of the accepting players from the final epoch to the Estimation Mechanism, is α-accurate.*

6.4 Benchmark Comparison

In this section we compare the cost of our mechanism to the cost of an omniscient mechanism that is constrained to make envy-free offers and achieve $\Theta(\alpha)$-accuracy. For the purposes of the cost comparison, in this section we assume that the individuals our algorithm approaches have linear cost functions: $c_i(\varepsilon) = v_i\varepsilon$ for some $v_i \in \mathbb{R}^+$.

Let $v(\alpha)$ be the smallest value v such that $\Pr_{x_i \sim \mathcal{D}}[v_i \leq v] \geq 1 - \alpha$. In other words, $(v(\alpha) \cdot 2\varepsilon, \varepsilon, \varepsilon)$ is the cheapest take-it-or-leave-it offer for ε-units of privacy that in the underlying population distribution would be accepted with probability at least $1 - \alpha$, It follows that:

Lemma 2. *Any $(\alpha/32)$-accurate mechanism that makes the same take-it-or-leave-it offer to every individual $x_i \sim \mathcal{D}$ must in expectation pay in total at least $\Theta(\frac{v(\frac{\alpha}{8})}{\alpha})$. Note that because here we assume cost functions are linear, this quantity is fixed independent of the number of agents the mechanism draws from $\mathcal{D}$.*

We now wish to bound the expected cost of our mechanism, and compare it to our benchmark cost, BenchMarkCost $= \Theta(\frac{v(\frac{\alpha}{8})}{\alpha})$.

Theorem 5. *The total expected cost of our mechanism is at most*

$$\begin{aligned}\mathbb{E}[\mathit{MechanismCost}] &= O\left(\log\log\left(\alpha \cdot v(\alpha/8)\right) \cdot \mathit{BenchMarkCost} + \frac{1}{\alpha^2}\right) \\ &= O\left(\log\log\left(\alpha^2 \cdot \mathit{BenchMarkCost}\right) \cdot \mathit{BenchMarkCost} + \frac{1}{\alpha^2}\right).\end{aligned}$$

Remark 3. Note that the additive $1/\alpha^2$ term is necessary only in the case in which $v(\alpha/8) \leq (1+\eta)/\alpha$: i.e., only in the case in which the *very first* offer will be accepted by a $1 - \alpha/8$ fraction of players with high probability. In this case, we have started off offering too much money, right off the bat. An additive term is necessary, intuitively, because we cannot compete with the benchmark cost in the case in which the benchmark cost is arbitrarily small.[4]

7 Discussion

In this paper, we have proposed a method for accurately estimating a statistic from a population that experiences cost as a function of their privacy loss. The statistics we consider here take the form of the expectation of some predicate over the population. We leave to future work the consideration of other, nonlinear, statistics. We have circumvented the impossibility result of [2] by using a mechanism empowered with the ability to approach individuals and make them take-it-or-leave-it offers (instead of relying on a direct revelation mechanism), and by relaxing the measure by which individuals experience privacy loss. Moving away from direct revelation mechanisms seems to us to be inevitable: if costs for privacy can be correlated with private data, then merely asking for individuals to report their costs is inevitably disclosive, for any reasonable measure of privacy. On the other hand, we do not claim that the model we use for how individuals experience cost as a function of privacy is "the" right one. Nevertheless, we have argued that some relaxation away from individuals experiencing privacy cost entirely as a function of the differential privacy parameter of the entire mechanism is inevitable (as made particularly clear in the setting of take-it-or-leave-it offers, in which individuals in this model would accept arbitrarily low offers). In particular, we believe that the style of survey mechanism presented in this paper, in which the mechanism may approach individuals with take-it-or-leave-it offers, is realistic, and any reasonable model for how individuals value their privacy should predict reasonable behavior in the face of such a mechanism.

Acknowledgements Work by K. Ligett was partially supported by NSF grant DMS-1004416. Work by A. Roth was partially supported by NSF grant CCF-1101389.

References

1. Heckman, J.: Sample selection bias as a specification error. Econometrica: Journal of the econometric society (1979) 153–161
2. Ghosh, A., Roth, A.: Selling privacy at auction. In: EC 2011: Proceedings of the 12th ACM conference on Electronic commerce, ACM (2011) 199–208
3. McSherry, F., Talwar, K.: Mechanism design via differential privacy. In: Proceedings of the 48th Annual Symposium on Foundations of Computer Science. (2007)
4. Chen, Y., Chong, S., Kash, I., Moran, T., Vadhan, S.: Truthful mechanisms for agents that value privacy. Arxiv preprint arXiv:1111.5472 (2011)

[4] We thank Lisa Fleischer and Yu-Han Lyu for pointing out the need for the additive term.

5. 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. Volume 3876 of Lecture Notes in Computer Science., Springer (2006) 265
6. Dwork, C.: Differential privacy: A survey of results. In: Proceedings of Theory and Applications of Models of Computation, 5th International Conference, TAMC 2008. Volume 4978 of Lecture Notes in Computer Science., Springer (2008) 1
7. Gupta, A., Ligett, K., McSherry, F., Roth, A., Talwar, K.: Differentially Private Combinatorial Optimization. In: Proceedings of the ACM-SIAM Symposium on Discrete Algorithms. (2010)
8. Nissim, K., Smorodinsky, R., Tennenholtz, M.: Approximately optimal mechanism design via differential privacy. In: ITCS 2012: Proceedings of the 3rd Innovations in Computers Science Conference. (2012)
9. Xiao, D.: Is privacy compatible with truthfulness? (2011) manuscript.
10. Nissim, K., Orlandi, C., Smorodinsky, R.: Privacy-aware mechanism design. In: ACM Conference on Electronic Commerce. (2012) 774–789
11. Huang, Z., Kannan, S.: The exponential mechanism for social welfare: Private, truthful, and nearly optimal. In: FOCS 2012. (2012)
12. Feigenbaum, J., Jaggard, A., Schapira, M.: Approximate privacy: foundations and quantification. In: Proceedings of the 11th ACM conference on Electronic commerce, ACM (2010) 167–178
13. Fleischer, L., Lyu, Y.H.: Approximately optimal auctions for selling privacy when costs are correlated with data. In: ACM Conference on Electronic Commerce. (2012) 568–585
14. Roth, A., Schoenebeck, G.: Conducting truthful surveys, cheaply. In: ACM Conference on Electronic Commerce. (2012) 826–843