Epsilon Voting: Mechanism Design for Parameter Selection in Differential Privacy

Nitin Kohli
School of Information
UC Berkeley
Email: nitin.kohli@ischool.berkeley.edu

Paul Laskowski
School of Information
UC Berkeley
Email: paul@ischool.berkeley.edu

***Abstract*—The behavior of a differentially private system is governed by a parameter epsilon which sets a balance between protecting the privacy of individuals and returning accurate results. While a system owner may use a number of heuristics to select epsilon, existing techniques may be unresponsive to the needs of the users who's data is at risk. A promising alternative is to allow users to express their preferences for epsilon. In a system we call epsilon voting, users report the parameter values they want to a *chooser mechanism*, which aggregates them into a single value. We apply techniques from mechanism design to ask whether such a chooser mechanism can itself be truthful, private, anonymous, and also responsive to users. Without imposing restrictions on user preferences, the only feasible mechanisms belong to a class we call randomized dictatorships with phantoms. This is a restrictive class in which at most one user has any effect on the chosen epsilon. On the other hand, when users exhibit single-peaked preferences, a broader class of mechanisms - ones that generalize the median and other order statistics - becomes possible.**

I. Introduction

When designing computer systems that handle personal information, special techniques are required to safeguard the privacy of vulnerable individuals. In recent years, differential privacy has emerged as the gold standard for privacy-protecting algorithms. Differential privacy is not a single technique, but rather a mathematical standard that quantifies the level of privacy protection [1]. In recent years, numerous protocols have been developed that meet this definition [2], [3].

Unfortunately, there is a cost to incorporating differential privacy into a computer system. Typically, the guarantee is achieved by combining computations with a precise amount of noise or randomness. As a consequence, the accuracy of a system tends to degrade as privacy protection increases. This trade-off is tunable by a parameter ϵ, which takes values between 0 and ∞. Conceptually, ϵ represents the maximum amount of information that an adversary can learn about an individual by observing the output.

The choice of ϵ matters. Although a company might advertise a system that is differentially private, if ϵ is set too high, the privacy guarantee is diluted and sensitive information is likely to be exposed. If ϵ is set too low, accuracy may degrade to the point that a system is not usable. Choosing an appropriate ϵ is therefore a critical step in deploying a differentially private system. Unfortunately, the definition of differential privacy offers little insight into how this should be done. Within the research community, there have been diverse proposals for choosing ϵ with no consensus, such as calibrating the choice of ϵ with threshold values of statistical estimators [4], examining adversarial attack models and potential queries of interest [5], and even building an economic model to compute the set of feasible ϵ in a world where individuals can be compensated to offset the risk incurred from participating in a study [6]. One simple idea that has not received much attention is to ask the people who's information is at risk how much privacy they want. No major deployment we are aware of has taken this tactic.

There are several reasons we may want to incorporate user preferences when choosing ϵ. First, company executives and data scientists may not understand the privacy risks that users face. This is particularly concerning to vulnerable and underrepresented subpopulations. Whether based on political viewpoint, religious conviction, or sexual orientation, minority groups may face disproportionate impacts in the case of a privacy breach. Complicating matters, many companies possess sensitive data on a global scale. A privacy decision made in a Silicon Valley board room may increase the risk to gay men living in Nigeria. Differential privacy has the potential to protect such groups, but only if ϵ is chosen with their concerns in mind.

A second reason to incorporate user preferences when choosing ϵ is to allow individuals to participate in the governance of systems that affect their lives. Numerous authors have argued that citizens should be allowed to make informed decisions about the the collection and use of their personal data [7]. Incorporating user preferences in the choice of ϵ achieves this by encoding these societal concerns about the use of personal information into the value of ϵ itself. Such inclusions of societal preferences, paired with the rigorous safeguards of differential privacy, may even encourage users to contribute information to public databases and other scientific endeavours.

Lastly, incorporating user preferences for ϵ can be a smart business strategy. In the modern economy, companies can differentiate themselves from their competitors by giving users input into what is currently a closed process. Such a tactic may serve as a signal to users that their concerns are taken seriously, while simultaneously protecting them from potential

harms that can arise from analysis of their information.

This study envisions a system in which users submit their personal preferences for the privacy parameter. That is, each user messages the system with the value that they would most want ϵ to be set to. It is the job of a *chooser mechanism* to aggregate these preferences in some way and output a final choice of ϵ that is used to protect user data. We will refer to such a system as ϵ voting. Voting may be restricted to only include the consumers using a system, or some votes may be reserved for a company or governing authority to give them more influence over the outcome.

It is important to stress that we're using the word voting in a general (game-theoretic) sense, not just to refer to majority voting. The chooser mechanism we envision could implement any map from incoming votes to its selected ϵ, and it need not be deterministic. In fact, we will soon provide an argument for why it is important to incorporate randomness into the chooser mechanism.

There are several important properties that we may want a chooser mechanism to have. We present them here in an intuitive form, reserving mathematical definitions for later sections. First, we want the mechanism to be *truthful*, meaning that users are not incentivized to misrepresent their preferences in hopes of getting a more preferred outcome. Truthful reporting ensures that we gain an accurate view of what users actually want. There is however a second, more important, reason to insist on truthful reporting. A classic result known as the revelation principle tells us that any outcome that can be achieved by a mechanism in general can also be achieved by a truthful mechanism. Because of this, we can restrict our analysis to truthful mechanisms without reducing the range of achievable outcomes, while also simplifying the analysis.

Second, the chooser mechanism must itself must keep individual votes private. To motivate this property, consider the example of an individual that is worried about an adversary learning her immigration status, which is listed in a database. An undocumented migrant might vote for a low ϵ to gain stronger privacy protection. As a result, across the population, votes will tend to correlate with immigration status. If an adversary is able to infer that an individual cast a vote for a small ϵ, they may further conclude that the individual is undocumented, circumventing the privacy guarantee that applies to the database. Because of this, it is vital for the chooser to be a privacy-preserving mechanism.

Third, we require that the chooser mechanism satisfy the game-theoretic notion of anonymity, which requires that the mechanism does not treat votes differently based on the identity of the voter. Suppose a voter Alice reports ϵ_A to the mechanism, while a voter Bob reports ϵ_B to the mechanism. Anonymity ensures that the behavior of the chooser mechanism does not change if Alice were to instead vote for ϵ_B and Bob were to vote for ϵ_A. Intuitively, anonymous mechanisms are not allowed to treat some voters as more important than others. While the term anonymity invokes numerous perceptions in the privacy community, we will use it in this narrow game-theoretic sense.

Finally, we would like our chooser mechanisms to be responsive to the votes of individuals. A number of classic impossibility results suggest that this is difficult to achieve. For example, the Gibbard-Satterthwaite Theorem tells us that if a chooser is deterministic and truthful, and there are more than two possible values of ϵ, it must be a dictatorship, meaning there is one individual who determines the outcome [8], [9]. However, such impossibility results hold for deterministic mechanisms, and can indeed be avoided by utilizing random mechanisms [10], [11]. Since differential privacy requires randomness, it may be possible to find private chooser mechanisms that are are not dictatorships.

This study therefore centers around a single research question:

> *What social outcomes can be achieved through an ϵ voting mechanism that is truthful, private, and anonymous?*

One contribution of our work is the introduction of truthful, anonymous, private mechanisms that can be used to aggregate user preferences for ϵ. We present such mechanisms for two settings. First, we consider the case that users have arbitrary preferences over ϵ. Next, we consider what happens when preferences are single-peaked. This is a property that is commonly studied in mechanism design and may realistically describe many privacy scenarios. The mechanisms we define can be applied to aggregate user privacy preferences in real world scenarios.

Where our work is most unique, however, is in the way we derive both both forward and reverse results. That is, instead of just providing examples of mechanisms, we will describe the set of *all* truthful, private, anonymous chooser mechanisms. We are able to do this for both the domain of unrestricted preferences and for single-peaked preferences. To the best of our knowledge, ours is the first result of this type for differentially private mechanisms.

The rest of the paper is organized as follows. Section 2 discusses the related work we draw upon. In Section 3, we formalize the structure of an ϵ voting system. In Section 4, we describe our model for users, which is based on mechanism design. In Section 5, we characterize the outcomes that may be implemented through truthful, private, and anonymous chooser mechanisms over an unrestricted domain of user preferences. In Section 6, we characterize outcomes for a class of user preferences known as single-peaked, which may be justified in certain privacy scenarios. We provide some discussion in Section 7.

II. Related Work

Our work draws on three main bodies of literature: differentially private mechanism design, epsilon selection techniques, and social choice theory. We discuss each in turn.

A. Truthful and Private Mechanisms

Differential privacy is introduced in Dwork et. al. [1] as a way to quantify the worst-case privacy loss that a data subject incurs from the operation of an algorithm. Generally,

differential privacy is achieved by masking the contribution of a single data point by utilizing noise during computations. This study also introduces the Laplace mechanism as a means to satisfy this notion for certain numeric queries by computing the true value of a query and then perturbing the result.

Shortly afterwards, McSherry and Talwar [12] introduce differential privacy as a solution concept in mechanism design, noting that it automatically achieves *approximate truthfulness*. In contrast to the standard notion of truthfulness, in which players have no incentive to strategically manipulate their preferences, approximate truthfulness requires that each player's incentive to misrepresent their type is bounded. McSherry and Talwar also introduce the exponential mechanism, a general framework that applies to both numeric and non-numeric outputs and functionally represents all differentially private mechanisms.

Given this framework, numerous papers have studied the relationship between truthfulness and differential privacy. For a comprehensive survey of this work, see Pai and Roth (2013) [13] and Dwork and Roth [14]. While we do not intend to reproduce the details of the survey here, we discuss three pieces of work that are relevant to our inquiry. The first is that of Huang and Kannan [15]. Using the exponential as a building block, Huang and Kannan construct a generalization of the VCG mechanism that was both private and truthful, assuming side payments are allowed. Nissim et. al. [16] address this question as well without the need for side payments in a non-standard mechanism design framework, by modeling player's utilities as not only functions over types and outcomes, but over reactions to those outcomes as well. In our setting, we prohibit side payments and work in the traditional mechanism design framework where utilities are over types and outcomes.

The study that is most closely related to our own is by Xiao [17]. In this work, Xiao constructs a method that transforms truthful and efficient mechanisms into one that preserves these properties while additionally satisfying *approximate* differentially private by sampling from a noisy histogram over the inputs. This technique of constructing a noisy histogram could be leveraged for our case, except that we require strict differential privacy and not the approximate version.

In contrast to the discussed works, our work contributes to the literature by fully characterizing the set of all truthful, private, and anonymous mechanisms for utility functions over two classes of important utility scales: the unrestricted domain and single-peaked preferences without the need for non-standard modeling frameworks, approximate privacy guarantees, or side payments.

B. Choosing Epsilon

Within the differential privacy literature, truthful mechanisms and methods for determining an epsilon have been addressed separately, but not together. As alluded to in the introduction, there have been many proposed methods for choosing an *optimal* ϵ. The diversity of solutions can be traced back to the different objectives that each study attempts to optimize. For example, Naldi and D'Acquisto [4] present a method for analytically computing the optimal value of ϵ to meet a certain level of accuracy for the Laplace mechanism. Lee and Clifton [5], on the other hand, analyze optimality with respect to an adversarial model and a class of potential queries of interest to an adversary. Hsu et. al. [6] take a very different approach to the previous two methods. In this case, optimality is determined by the criteria of incentivizing a data subject to opt-in to a study that would otherwise not take part in. Our work addresses the question of choosing ϵ according to individual preferences.

C. Social Choice Theory

To investigate what ϵ can be chosen for particular preferences, we draw upon social choice theory. As a branch of mechanism design, social choice theory is interested in determining how to consolidate a collection of individual preferences into a single preference that is reflective of the society as a whole. One particular application of social choice theory is voting. In this case, questions lie around the existence and construction of voting mechanisms that satisfy certain criteria. However, it often turns out that, depending on the criteria in question, no such mechanism may exist.

The Gibbard-Satterthwaite [8], [9] Theorem states that any deterministic and truthful mechanism that acts over a ballot with at least three candidates where voters can have arbitrary preferences over the candidates must be a dictatorship. Later works demonstrated ways to circumvent this dictatorial result. Gibbard [10], [11] shows that it is possible to avoid dictatorships by incorporating voting mechanisms that utilized randomness. Additionally, Moulin [18] demonstrates that this dictatorial result can be avoided under a more restrictive set of assumptions, even with deterministic mechanisms. Instead of allowing voters to have arbitrary preferences over a discrete set of options, Moulin considers voters with single-peaked preferences over the real line $\mathbb{R}$. In this setting, the only deterministic mechanisms that are simultaneously truthful, anonymous, and range-preserving are the median and other order statistics. The contributions of our work can be viewed as differentially private extensions of these results to the problem of choosing an ϵ.

III. ϵ Voting Framework

Our framework centers around a chooser mechanism C, which has the purpose of selecting a single parameter ϵ. We imagine a set of n users, each with preferences over possible parameter values. We consider a discrete set of values called the ballot $B = \{\epsilon_1, ..., \epsilon_k\}$, where $k \geq 2$. Without loss of generality, we assume $0 < \epsilon_1 < ... < \epsilon_k < \infty$.

In practice, one could imagine an ϵ voting system with more general ballot entries. For example, instead of having users select ϵ directly, users could instead choose from a collection of candidate values that encode ϵ in a more comprehensible way. For the purpose of our analysis, we restrict ourselves to the case of choosing ϵ directly.

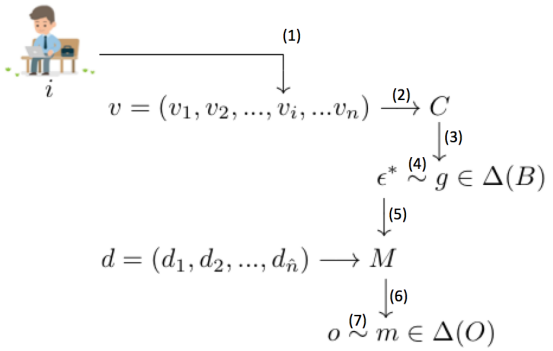

Fig. 1. A chooser mechanism combines votes and selects a single ϵ^*.

A. Voting Process

The operation of our chooser mechanism is shown in Figure 1. Starting at the top of this figure, each user i selects a preferred value from the ballot, $v_i \in B$, which we call user i's vote (1). Let $v = (v_1, v_2, ..., v_n)$ be the vector of all votes. This vector is supplied as an input to chooser C (2).

The chooser may incorporate randomness into its output. For a given input v, there is some probability that C selects each ϵ in B. Let $\Delta(B)$ be the set of all probability distributions over B. Given an input v, we can view C as selecting a single probability distribution from this set, $g(v) \in \Delta(B)$ (3). We will use subscripts to denote the components of a probability distribution; for $z \in B$, $g(v)_z$ is the probability that $g(v)$ assigns to z. The specific choice of ϵ^* is a draw from distribution g (4).

ϵ^* is used as an input to the main mechanism M (5) (possibly scaled, as we explain below). Mechanism M operates on a database d consisting of entries $d_1, d_2, ..., d_{\hat{n}}$. Often, we will assume that each user is associated with exactly one database entry, but that may not always be the case. For example, in a database of rides that the users have taken on a transit service, we could identify the entries of d with individual rides or with individual users.

Much like C, M may incorporate randomness into its output. We let O be the set of possible outputs of M, and $\Delta(O)$ be the set of probability distributions over this set. We view M as outputting a single distribution, $m(d) \in \Delta(O)$ (6). The output o may be seen as a draw from this distribution (7).

B. Privacy and Anonymity

We review the definition of differential privacy in the context of mechanism M. Conceptually, a mechanism is differentially private if the probability of any output from the mechanism doesn't change "too much" based on a single data point.

Definition 1. *Mechanism M satisfies ϵ differential privacy if, for all databases d and $\hat{d}$ that differ on at most one element, and for all measurable sets $S \subseteq O$,*

$$\mathbb{P}_{m(d)}(S) \leq e^{\epsilon} \mathbb{P}_{m(\hat{d})}(S)$$

where $\mathbb{P}_x(S)$ is the probability of S under distribution x.

As we mentioned earlier, it is important that C itself be a private mechanism. A user's vote for ϵ may reveal information about the contents of database d. After all, asking for strong privacy may be taken as a signal that a user has something to hide. If an adversary can learn the votes of individuals, they may therefore circumvent any privacy guarantee on M.

One idea is to choose a separate ϵ_C before any votes are cast and require that C satisfy ϵ_C differential privacy. The definition mirrors the one above, but the fact that B is finite simplifies notation.

Definition 2. *Mechanism C satisfies ϵ_C differential privacy if, for all vote vectors v and $\hat{v}$ that differ on at most one element, and for all $z \in B$,*

$$g(v)_z \leq e^{\epsilon_C} g(\hat{v})_z$$

This approach is simple but has some disadvantages. If ϵ_C is set too high and the chooser outputs some far smaller ϵ^*, the privacy guarantee of the main mechanism may be meaningless since more information is being leaked by the chooser than the main mechanism. On the other hand, it may be wasteful to set ϵ_C smaller than needed, since this will result in a lot of noise when choosing ϵ^*.

Another approach is not to set C's privacy parameter in advance. Instead, consider setting aside a fraction $\lambda \in (0, 1)$ of the computed privacy budget ϵ^* to apply back to C.

To motivate this idea, suppose that we knew in advance the exact ϵ^* the chooser would select to be the total privacy budget. Since we require both C and M to be privacy preserving, we need to allocate part of this budget to C and the remaining part to M. That is, we could set aside some fraction $\lambda\epsilon^*$ of the budget to be used by C to compute ϵ^*, and leave the remaining $(1 - \lambda)\epsilon^*$ to be used in M. Since the sequential application of differentially private mechanisms results in additive cumulative privacy loss [14], constructing C and M in such a manner utilizes the entire privacy budget ϵ^*.

However, we do not know the value of ϵ^* prior to running C. To circumvent this issue, we adapt the definition of differential privacy by writing the privacy bound for each output $z \in B$ as though that were the selected ϵ.

Definition 3. *Take $\lambda \in (0, 1)$. A mechanism $C : B^n \to \Delta(B)$ satisfies (λ, B) differential privacy if, for all vote vectors v and $\hat{v}$ that differ on at most one element, and for all $z \in B$,*

$$g(v)_z \leq e^{\lambda z} g(\hat{v})_z$$

In other words, we require C to circularly satisfy a differential privacy guarantee over the candidate it would choose. Note that the definition is the same as above, except ϵ_C has been replaced by λz. Loosely speaking, there is a differential privacy bound at each $z \in B$, but the bound is stricter or looser depending on the z in question.

To further motivate this approach, consider what happens when the chooser outputs a specific ϵ^*. If an attacker observes the outputted ϵ^*, then the likelihood ratio between v and $\hat{v}$, $g(\hat{v})_z / g(v)_z$, is bounded by $\lambda\epsilon^*$. This bound is identical to the one that would result for $\lambda\epsilon^*$ differential privacy.

Practically speaking, this suggests that the chooser must use more noise when outputting a smaller ϵ^*, but may be more accurate when outputting a higher ϵ^*. This approach can therefore make more efficient use of the available privacy budget while protecting users. For the rest of this paper, we will use (λ, B) differential privacy. However, it is straightforward to adapt all of the results to accommodate ϵ_C differential privacy.

In addition to (λ, B) differential privacy, we also require our mechanisms to satisfy anonymity. Conceptually this criteria states that the mechanism should not discriminate based on who is submitting a particular vote.

Definition 4. *A mechanism g is anonymous if, for any permutation $\pi : B^n \to B^n$, and for any $v \in B^n$, $g(v) = g(\pi(v))$.*

C. Example Chooser Mechanisms

In this section, we provide example mechanisms to motivate our theoretical results in Sections 5 and 6. We begin with a mechanism that is entirely unresponsive to voters, but will be a useful building block for other mechanisms.

Example 1. A mechanism g is *degenerate* if for every pair of vote vectors $v, w \in B^n$, $g(v) = g(w)$.

Constructing degenerate mechanisms is straightforward: take any $b = (b_1, ..., b_k) \in \Delta(B)$ and suppose $g(v) = b$ for all vote vectors v. Such mechanisms trivially satisfy (λ, B) differential privacy and anonymity, but are completely non-responsive to individuals preferences. We will refer to mechanisms that are not degenerate as *non-degenerate*.

Given a player i, we denote an *environment* v_{-i} as the votes that were cast by all players except i [11]. We also use (z, v_{-i}) to refer to the votes when player i votes for z and all others cast votes v_{-i}, namely $(v_1, ..., v_{i-1}, z, v_{i+1}, ..., v_n)$.

Denote the lottery that outputs $z \in B$ with probability 1 as δ^z. In other words, the x^{th} component of δ^z is $\delta^z_x = \mathbb{I}(x = z)$, where $\mathbb{I}$ represents the indicator function.

Example 2. A mechanism g is a *deterministic dictatorship* if there exists some player $i \in [n]$ such that for all $v_{-i} \in B^{n-1}$ and for all $z \in B$, $g(z, v_{-i}) = \delta^z$.

Deterministic dictatorships emerge frequently in the mechanism design literature, but they cannot satisfy (λ, B) differential privacy as there is no randomness involved in the mechanism. Furthermore, in terms of responsiveness, only a single voter has any influence over the outcome of the mechanism. Such a mechanism is primarily interesting as an extreme case, and as a component of more complex mechanisms.

Denote the number of votes that alternative z receives in v as $n_z(v) = |\{i \in [n] | v_i = z\}|$.

Example 3. A mechanism g is called a *randomized dictatorship* if

$$g(v)_z = \frac{n_z(v)}{n}$$

Note that g is indeed a valid probability distribution as $n = \sum_{z \in B} n_z(v)$.

In this mechanism, the probability of any ϵ is proportional to the number of votes for that ϵ. It can be understood in terms of the following procedure: Select a single user at random, then do exactly what that user wants. This is mathematically equivalent to a probability mixture of deterministic dictatorships. A randomized dictatorship is anonymous, but it is explicitly not private, as the following proposition shows.

Proposition 1. *If g is a randomized dictatorship, then g cannot be (λ, B) differentially private.*

Proof. By way of contradiction suppose g is (λ, B) differentially private. Then for all vote vectors $v, \hat{v} \in B^n$ that differ by at most one vote, and for all $y \in B$, $g(\hat{v})_y \leq e^{\lambda y} g(v)_y$. Take any $x, y \in B$ distinct and define $v_{-i} = (x, ..., x)$, $v = (x, v_{-i})$, and $\hat{v} = (y, v_{-i})$. Then,

$$0 < \frac{1}{n} = g(\hat{v})_y \leq e^{\lambda y} g(v)_y = 0$$

yielding a contradiction. □

Thus, we see that our requirement for privacy automatically rules out such dictatorial behavior. Intuitively, randomized dictatorships fail to meet our privacy requirement since there is no "plausible deniability" over outcomes that receive just a single vote. One idea to remedy the situation is to add some "buffer" to each ϵ so that the associated probability doesn't get too close to zero.

Example 4. A mechanism g is a *randomized dictatorship with phantoms* if

$$g(v)_z = \frac{n_z(v) + \phi_z}{n + \sum_{x \in B} \phi_x}$$

with $\phi_x > 0$ for all $x \in B$.

We refer to each ϕ_z as the phantom votes for z. While the use of the term phantom votes may suggest that these ϕ_z take integer values, they are in fact more general and can take on any positive non-integer value as well. This mechanism behaves similarly to the randomized dictatorship, with the exception that their always exists some non-zero probability of any $z \in B$ being output by the mechanism.

Note that this mechanism can be constructed as a probability mixture of the randomized dictatorship and a degenerate mechanism which assigns probability $\phi_z / \sum_{x \in B} \phi_x$ to each $z \in B$. We will explore the game theoretic and privacy properties of this mechanism in Section 5.

Our final example chooser is motivated by Moulin's use of order statistics in the deterministic setting [18]. For a vote vector v, let $R_z(v)$ be the set of *ranks* of the votes cast for z. That is, let v' be a permutation of v such for $i > j$, $v'_i \geq v'_j$. Then $R_z(v) = \{i \in [n] | v'_i = z\}$.

Example 5. We define a *randomized median* mechanism as follows. Define probabilities $\xi_1, \xi_2, ..., \xi_k, t(1), t(2), ..., t(n)$ that collectively sum to 1. Define

$$g(v)_z = \xi_z + \sum_{r \in R_z(v)} t(r)$$

As a specific case, when n is odd take $t(r) = \mathbb{I}(r = \lceil \frac{n}{2} \rceil)$ and all $\xi_z = 0$. This results in the familiar median. Other order statistics can be created in the same way. In general, a randomized median can be viewed as a probability mixture of order statistics, and a denegerate distribution comprised of $\xi_1, \xi_2, ..., \xi_k$.

As a final special case, consider $t(r) = t(r')$ for all $r, r' \in [n]$. This results in a randomized dictatorship with phantoms. We will discuss the randomized median mechanism further in Section 6.

D. Algebraic Properties of Vote Switching

At a fundamental level, (λ, B) differential privacy is concerned with vectors that differ in the vote of at most one player. We are therefore interested in describing the what happens when a single player switches their vote. That is, if a player were to change her vote from x to y, how does the new lottery $g(y, v_{-i})$ compare with the previous one $g(x, v_{-i})$? As observed by Gibbard (1978) [11], the additive effect of switches plays a pivotal role in constraining the behavior of possible mechanisms over a given utility domain. To formalize this, we define a *switch* and its corresponding *effect* as follows.

Definition 5. *Take $v \in B^n$. A switch by player i from candidate x to y is a function $S^i_{x,y} : B^{i-1} \times \{x\} \times B^{n-i} \to B^{i-1} \times \{y\} \times B^{n-i}$ such that $S^i_{x,y}(x, v_{-i}) = (y, v_{-i})$ for any environment v_{-i}.*

Definition 6. *Take $v \in B^n$. The effect of a switch $S^i_{x,y}$ on vote vector v with $v_i = x$ is*

$$A(S^i_{x,y}, v) = g(S^i_{x,y}(v)) - g(v).$$

If g is an anonymous mechanism, the effect of a switch does not depend on which player switches their vote. We will assume this is true for the remainder of the paper, and therefore omit the superscript that indicates a specific player.

The following lemmas describe some basic properties of switches. First off, switches commute with each other.

Lemma 1. *Fix $v \in B^n$. For all $a, b, x, y \in B$, $S_{x,y} \circ S_{a,b} = S_{a,b} \circ S_{x,y}$.*

Proof. Take $v \in B^n$ with $v_i = a$ and $v_j = x$. Observe that $S_{x,y} \circ S_{a,b}(v) = S_{x,y}(b, v_{-i})$. Furthermore, $S_{x,y}(b, v_{-i}) = (y, b, v_{-\{i,j\}})$, where $v_{-\{i,j\}}$ denotes the votes of all voters except i and j. This is the same as $S_{a,b}(y, v_{-j})$, which is indeed $S_{a,b} \circ S_{x,y}(v)$. □

In addition, the effect of composed switches can be additively decomposed.

Lemma 2. *Fix $v \in B^n$. For all $a, b, x, y \in B$, $A(S_{x,y} \circ S_{a,b}, v) = A(S_{a,b}, v) + A(S_{x,y}, S_{a,b}(v))$.*

Proof. Take $v \in B^n$ with at least one vote for a and at least one vote for x. Observe that $A(S_{x,y} \circ S_{a,b}, v) = g\big(S_{x,y}(S_{a,b}(v))\big) - g(v)$. Adding and subtracting $g(S_{a,b}(v))$ from the right hand side and grouping terms produces $A(S_{x,y}, S_{a,b}(v)) + A(S_{a,b}, v)$. □

IV. User Model and Truthfulness

Next, we formalize the user dynamics of this framework using mechanism design. Suppose each user has a global utility function $U_i : B \times O \to \mathbb{R}$. For all $i \in [n]$, user i's privacy-specific utility function is given by

$$u_i(\epsilon; C) = \mathbb{E}_{o \sim M_\epsilon}[U_i(\epsilon, o; C)]$$

Conceptually, $u_i(\epsilon; C)$ denotes player i's expected global utility when ϵ is used to protect their privacy, taken over the randomness of all possible outcomes of M. In principle, utility for a given ϵ and o could depend on the chooser C. This is because depending on the nature of C, a particular ϵ may leak more or less information about the preferences of users. For the rest of this paper, however, our results will refer to a fixed C, so we generally refer to this quantity as $u_i(\epsilon)$. Since B is finite, we can represent the utility over all candidates in the ballot with a single vector $u_i \in \mathbb{R}^k$, where the j^{th} component of this vector refers to $u_i(\epsilon_j)$.

Given a ballot B, each voter i has some preference ordering over B. We will refer to the top candidate in a player i's preference profile as their *type*, which we denote as α_i. More formally, α_i is a type for voter i if $u_i(\alpha_i) \geq u_i(\beta)$ for all $\beta \in B$. We say that a voter is *indifferent* between candidates x and y if $u_i(x) = u_i(y)$. If a voter is indifferent between their top candidates over B, we will refer to each of these top candidates as player i's *types*.

As is common in the mechanism design literature, we assume that voters seek to optimize their expected utility under the mechanism. Formally, suppose voter i reports $v_i \in B$ to the mechanism. Then the expected utility for player i, when other voters submit $v_{-i} \in B^{n-1}$, is given by the dot product:

$$\mathbb{E}_{\epsilon \sim g(v_i, v_{-i})}[u_i(\epsilon)] = u_i \cdot g(v_i, v_{-i})$$

Using this machinery, we are ready to formalize truthfulness. Intuitively, a mechanism g is truthful if truthfully reporting a type to the mechanism is at least as good as lying.

Definition 7. *A mechanism g is truthful if reporting a type is an undominated strategy. That is, for all voters i with type α_i, for all $\beta \in B$, and for all $v_{-i} \in B^{n-1}$,*

$$u_i \cdot g(\alpha_i, v_{-i}) \geq u_i \cdot g(\beta, v_{-i})$$

Notice that when a player has multiple types, reporting any one of them meets the requirement for truthfulness. In this case, switching between the utility-maximizing alternatives does not alter the expected utility of the player. This is formalized by the following lemma.

Lemma 3. *Suppose g is truthful. If $x, y \in \text{argmax}_{z \in B} u_i(z)$, then for any environment $v_{-i} \in B^{n-1}$, $u_i \cdot A(S_{x,y}, (x, v_{-i})) = 0 = u_i \cdot A(S_{y,x}, (y, v_{-i}))$.*

Proof. Since both $u_i(x) \geq u_i(z)$ and $u_i(y) \geq u_i(z)$ for all $z \in B$, $u_i(x) = u_i(y)$. The truthfulness of g implies that $u_i \cdot g(x, v_{-i}) \geq u_i \cdot g(y, v_{-i})$ and $u_i \cdot g(y, v_{-i}) \geq u_i \cdot g(x, v_{-i})$, establishing $u_i \cdot g(y, v_{-i}) = u_i \cdot g(x, v_{-i})$. Algebraic manipulation yields $u_i \cdot \big(g(y, v_{-i}) - g(x, v_{-i})\big) =$

$0 = u_i \cdot \big(g(x, v_{-i}) - g(y, v_{-i})\big)$. Applying the definition of effect to both sides proves the claim. □

V. Mechanisms Over the Unrestricted Domain

In this section, we explore mechanisms over the unrestricted domain. That is, we allow for utility functions to attain arbitrary values without any restriction on their functional form. We begin with a characterization of the entire class of mechanisms that satisfy our desiderata.

Theorem 1. *Suppose $|B| \geq 3$. Over the unrestricted domain, g is truthful, (λ, B) differentially private, anonymous, and non-degenerate if and only if g is a randomized dictatorship with phantoms*

$$g(v)_z = \frac{n_z(v) + \phi_z}{n + \sum_{x \in B} \phi_x}$$

where $\phi_x \geq \big(e^{\lambda x} - 1\big)^{-1}$ for all $x \in B$.

Proof. ($\Longleftarrow$) Let $M = n + \sum_{x \in B} \phi_x$. Suppose $g(v)_z = M^{-1}\big(n_z(v) + \phi_z\big)$ for constants $\phi_x \geq \big(e^{\lambda x} - 1\big)^{-1}$ for all $x \in B$. It is clear that g is anonymous and non-degenerate. To demonstrate the truthfulness of g, consider a voter i with type α_i. For any alternative $x \in B$, and for any $v_{-i} \in B^{n-1}$, observe that

$$\begin{aligned} & u_i \cdot \big(g(\alpha_i, v_{-i}) - g(x, v_{-i})\big) \\ &= \sum_{z \in B} u_i(z)\big(g(\alpha_i, v_{-i})_z - g(x, v_{-i})_z\big) \\ &= M^{-1} \sum_{z \in B} u_i(z)\big(n_z(\alpha_i, v_{-i}) - n_z(x, v_{-i})\big) \end{aligned}$$

Within this summand, $n_z(\alpha_i, v_{-i}) = n_z(x, v_{-i})$ for all $z \notin \{\alpha_i, x\}$. Furthermore, $n_x(\alpha_i, v_{-i}) - n_x(x, v_{-i}) = -1$ and $n_{\alpha_i}(\alpha_i, v_{-i}) - n_{\alpha_i}(x, v_{-i}) = 1$. As such,

$$u_i \cdot \big(g(\alpha_i, v_{-i}) - g(x, v_{-i})\big) = M^{-1}\big(u_i(\alpha_i) - u_i(x)\big) \geq 0$$

where the inequality follows due to $M > 0$ and α_i being voter i's type.

Lastly, we demonstrate that g satisfies (λ, B) differential privacy. Take any vote vectors (y, v_{-i}) and (z, v_{-i}) and consider the following ratio R_x:

$$\begin{aligned} R_x = \frac{g(v_i = y, v_{-i})_x}{g(v_i = z, v_{-i})_x} &= \frac{n_x(y, v_{-i}) + \phi_x}{n_x(z, v_{-i}) + \phi_x} \\ &= \frac{n_x(v_{-i}) + \mathbb{I}(y = x) + \phi_x}{n_x(v_{-i}) + \mathbb{I}(z = x) + \phi_x} \end{aligned}$$

Note that this ratio is maximized when $y = x$ and $z \neq x$. Since $\phi_x \geq \big(e^{\lambda x} - 1\big)^{-1}$, it follows that

$$R_x \leq \frac{n_x(v_{-i}) + 1 + \phi_x}{n_x(v_{-i}) + \phi_x} = 1 + \frac{1}{n_x(v_{-i}) + \phi_x} \leq 1 + \frac{1}{\phi_x} \leq e^{\lambda x}$$

establishing (λ, B) differential privacy.
($\Longrightarrow$) The details of this direction of the proof can be found in Appendix A. □

This theorem fully characterizes the set of non-degenerate, truthful, anonymous, (λ, B) differentially private mechanisms. There are a few key observations to make here. First off, note that $\phi_x \geq \big(e^{\lambda x} - 1\big)^{-1}$ implies $\phi_x > 0$. As such, the presence of phantoms is necessary to assure that g can meet the desired privacy guarantee.

Next, observe that the lower bound on ϕ_x is decreasing in x. Conceptually this means that more phantom noise is required to protect the privacy of those who voted for small x. This matches intuition, as more noise is indeed required to protect those who desire stronger privacy guarantees. Alternatively, when x is large this tells us that we require less of a contribution from ϕ_x, as these individuals who cast their vote for large x prefer strong accuracy guarantees over privacy.

In conclusion, the *only* mechanisms that satisfy our desiderata over the unrestricted domain are random dictatorship with phantoms. So while we technically met our goal of producing neither a degenerate nor dictatorial outcome, we ended up showing that the only ϵ voting mechanism without utility restrictions must be some probability mixture of the two.

VI. Mechanisms Over Single-Peaked Preferences

The analysis from the previous section demonstrates just how limited the class of truthful, private, and anonymous mechanisms is when operating in the unrestricted domain. Conceptually this makes sense, as both the space of truthful mechanisms and the space of private mechanisms are limited. Taking their intersection necessarily results in a space that is even smaller. In this section, we move away from the unrestricted domain setting by imposing an assumption called *single-peaked preferences* (SPP). Although this reduces generality, this is a common assumption that may be realistically applied to many scenarios.

Definition 8. *A utility function satisfies single-peaked preferences (SPP) over B if there exists some $\alpha \in B$, which we call a peak, such that for all $x, y \in B$, $x \leq y \leq \alpha \implies u(x) \leq u(y) \leq u(\alpha)$ and $\alpha \leq x \leq y \implies u(\alpha) \geq u(x) \geq u(y)$.*

A careful reader will note that this definition slightly differs from the definition of SPP presented in the seminal paper by Moulin [18]. In particular, Moulin's definition requires the peak be unique. In our setting, it is more natural to allow users to be indifferent among multiple top ϵ's.

SPP are well-motivated to capture many common privacy scenarios. We discuss a few of them here. First off, we can imagine a user who is very privacy conscious and does not care much for accurate analysis from the use of their data. In this case, the user's utility function would decrease as ϵ increases, satisfying SPP with $\alpha = \epsilon_1$. On the other end of the spectrum, a user may be solely interested in the furthering science without any regard for privacy at all. In this case, the user's utility function would be some increasing function of ϵ which satisfies SPP with $\alpha = \epsilon_k$.

SPP can also capture cases in between these two extremes. A user may value the insights that come from more accurate outputs up to a certain threshold, but consider further increases in ϵ not worth the privacy loss. In such scenarios, $\alpha \in \{\epsilon_2, ..., \epsilon_{k-1}\}$. It is possible to derive utility functions like

this under reasonably natural model assumptions. For example, several studies have encountered a $O(\epsilon^2)$ bound on the utility loss associated with privacy [19], [20]. This is suggestive of utility functions that are concave, at least for small ϵ.

Comparing to Moulin's results for the deterministic domain, one might wonder if the randomized median game is a truthful and anonymous mechanism. It turns out that this is correct, but it is not the only truthful and anonymous mechanism in this setting. Instead, we define a generalized version of this below.

Before we begin, we introduce new notation and terminology. For $z \in B$, denote the neighbors of z as z^+ and z^-. In particular, z^+ is the smallest $\epsilon \in B$ that is larger than z. Similarly, z^- is the largest $\epsilon \in B$ that is smaller than z. Such notation will be helpful when considering *neighboring switches*. Define the rank of z as the number of votes that vote for an alternative no larger than z, $r_z(v) = |\{i \in [n] | v_i \leq z\}|$.

Definition 9. *A mechanism g is a generalized randomized median if there exists a set of functions, which we call transfers, $\{t_z : [n] \to [0,1] | z \in B\}$ and a set of base masses $\{\xi_z \geq 0 | z \in B\}$ such that*

$$g(v)_z = \sum_{r=1}^{r_z(v)} t_z(r) - \sum_{r=1}^{r_{z^-}(v)} t_{z^-}(r) + \xi_z$$

holds for all $z \in B$ and is a valid probability distribution.

Notice that if we require $t_z(r) = t_{z'}(r)$ for all z, z', we recover the randomized median mechanism from Example 5.

Theorem 2. *For SPP, a mechanism g is truthful and anonymous if and only if g is a generalized randomized median.*

Proof. ($\Longleftarrow$) Anonymity is clear. To demonstrate truthfulness, it is sufficient to show that for a player i with type α_i and for all $v_{-i} \in B^{n-1}$, (1) $z \leq z^+ \leq \alpha_i$ implies $u_i \cdot g(z, v_{-i}) \leq u_i \cdot g(z^+, v_{-i})$, and (2) $\alpha_i \leq z \leq z^+$ implies $u_i \cdot g(z, v_{-i}) \geq u_i \cdot g(z^+, v_{-i})$.

To do so, we first show that $A(S_{z,z^+}, v)_y = 0$ if $y < z$ or $y > z^+$. Observe that if $y < z$, then $r_y(z, v_{-i}) = r_y(z^+, v_{-i})$ and $r_{y^-}(z, v_{-i}) = r_{y^-}(z^+, v_{-i})$. A similar argument holds for the case when $y > z^+$ as well. Using this insight about the ranks, it follows that

$$\begin{aligned}
&A(S_{z,z^+}, v)_y = g(S_{z,z^+}(v))_y - g(v)_y \\
&= g(z^+, v_{-i})_y - g(z, v_{-i})_y \\
&= \left(\sum_{r=1}^{r_y(z^+, v_{-i})} t_y(r) - \sum_{r=1}^{r_{y_-}(z^+, v_{-i})} t_{y_-}(r) \right) \\
&- \left(\sum_{r=1}^{r_y(z, v_{-i})} t_y(r) - \sum_{r=1}^{r_{y_-}(z, v_{-i})} t_{y_-}(r) \right) = 0
\end{aligned}$$

Next, we prove (1). Suppose that $z \leq z^+ \leq \alpha_i$. Then

$$\begin{aligned}
&u_i \cdot \big(g(z^+, v_{-i}) - g(z, v_{-i})\big) \\
&= \sum_{y \in B} u_i(y) \big(g(z^+, v_{-i})_y - g(z, v_{-i})_y\big) \\
&= \big(u_i(z^+) - u_i(z)\big) \big(g(z^+, v_{-i})_{z^+} - g(z, v_{-i})_{z^+}\big)
\end{aligned}$$

By SPP, $u_i(z^+) - u_i(z) \geq 0$. Also,

$$\begin{aligned}
&g(z^+, v_{-i})_{z^+} - g(z, v_{-i})_{z^+} \\
&= \left(\sum_{r=1}^{r_{z^+}(z^+, v_{-i})} t_{z^+}(r) - \sum_{r=1}^{r_{z^+}(z, v_{-i})} t_{z^+}(r) \right) \\
&+ \left(\sum_{r=1}^{r_z(z, v_{-i})} t_z(r) - \sum_{r=1}^{r_z(z^+, v_{-i})} t_z(r) \right)
\end{aligned}$$

Observe that $r_{z^+}(z^+, v_{-i}) = r_{z^+}(z, v_{-i})$, producing

$$g(z^+, v_{-i})_{z^+} - g(z, v_{-i})_{z^+} = \sum_{r=1}^{r_z(z, v_{-i})} t_z(r) - \sum_{r=1}^{r_z(z^+, v_{-i})} t_z(r)$$

Furthermore, either $r_z(z^+, v_{-i}) = r_z(z, v_{-i})$ or $r_z(z^+, v_{-i}) + 1 = r_z(z, v_{-i})$. In the first case, the difference in the sums is 0. In the second case, the difference in the sums is precisely $t_z(r_z(z, v_{-i})) \geq 0$. Thus, $g(z^+, v_{-i})_{z^+} - g(z, v_{-i})_{z^+} \geq 0$.

This yields $u_i \cdot \big(g(z^+, v_{-i}) - g(z, v_{-i})\big) \geq 0$, which concludes the proof of (1). By a symmetric argument it is easily seen that (2) holds as well, establishing truthfulness.

($\Longrightarrow$) The details of this direction of the proof can be found in Appendix B. □

Note that this result holds for any ballot with $|B| \geq 2$. Our result from the previous section over the unrestricted domain required $|B| \geq 3$. When $|B| = 2$, every utility function in the unrestricted domain satisfies our extended definition of SPP. This produces the following.

Corollary 1. *Suppose $|B| = 2$. Over the unrestricted domain, g is truthful and anonymous if and only if g is a generalized randomized median.*

Next, we turn our attention to (λ, B) differential privacy.

Lemma 4. *A generalized randomized median g meets the bound,*

$$g(v)_x / g(S_{z,z^+}(v))_x \leq e^{\lambda x}$$

for all $v \in B^n$, for all $x \in B$, and for all neighboring switches S_{z,z^+}, if and only if g is (λ, B) differentially private.

Proof. The reverse direction proceeds immediately from the definition of (λ, B) differential privacy. For the forward direction, consider a player i, an environment v_{-i}, and two votes, $\epsilon_a, \epsilon_b \in B$. Consider the case that $\epsilon_a < \epsilon_b$. We need to show that

$$g(\epsilon_a, v_{-i})_x / g(\epsilon_b, v_{-i})_x \leq e^{\lambda x},$$

for all $x \in B$. We can write the left hand side as the product

$$\frac{g(\epsilon_a, v_{-i})_x}{g(\epsilon_b, v_{-i})_x} = \frac{g(\epsilon_a, v_{-i})_x}{g(\epsilon_{a+1}, v_{-i})_x} \cdot \frac{g(\epsilon_{a+1}, v_{-i})_x}{g(\epsilon_{a+2}, v_{-i})_x} \cdot \ldots \cdot \frac{g(\epsilon_{b-1}, v_{-i})_x}{g(\epsilon_b, v_{-i})_x}$$

Notice that each term in the product has the form,

$$\frac{g(\hat{v})_x}{g(S_{z,z^+}(\hat{v}))_x}$$

for some vector $\hat{v}$ and some neighboring switch S_{z,z^+}. If $x \notin \{z, z^+\}$, the fraction is 1. If $x = z^+$, the fraction is at most

1 since the switch transfers probability into x, making the denominator larger than the numerator. If $x = z$, the fraction is at most $e^{\lambda x}$ by assumption. The entire product is therefore bounded by $e^{\lambda x}$ as required. A similar argument holds for the case that $\epsilon_b < \epsilon_a$, completing the proof. □

Theorem 3. *To satisfy (λ, B) differential privacy, the following system of linear constraints must hold for all $z \in B$ and all ranks ρ. For every $z > \epsilon_1$,*

$$\frac{t_z(\rho) + \sum_{r=1}^{\rho-1} \big(t_z(r) - t_{z^-}(r)\big) + \xi_z}{\sum_{r=1}^{\rho^-} \big(t_z(r) - t_{z^-}(r)\big) + \xi_z} \leq e^{\lambda z}$$

Additionally, for every $z < \epsilon_k$,

$$\frac{t_{z^-}(\rho) + \sum_{r=1}^{\rho} \big(t_z(r) - t_{z^-}(r)\big) + \xi_z}{\sum_{r=1}^{\rho} \big(t_z(r) - t_{z^-}(r)\big) + \xi_z} \leq e^{\lambda z}$$

Also,

$$\frac{t_{\epsilon_1}(\rho) + \sum_{r=1}^{\rho-1} t_{\epsilon_1}(r) + \xi_{\epsilon_1}}{\sum_{r=1}^{\rho-1} t_{\epsilon_1}(r) + \xi_{\epsilon_1}} \leq e^{\lambda \epsilon_1}$$

and

$$\frac{t_{\epsilon_{k-1}}(\rho) + \sum_{r=1}^{n} t_{\epsilon_k}(r) - \sum_{r=1}^{\rho} t_{\epsilon_{k-1}}(r) + \xi_{\epsilon_k}}{\sum_{r=1}^{n} t_{\epsilon_k}(r) - \sum_{r=1}^{\rho} t_{\epsilon_{k-1}}(r) + \xi_{\epsilon_k}} \leq e^{\lambda \epsilon_k}$$

Proof. We provide a computation for the first part of the claim, noting that the rest are easily computable. To satisfy (λ, B) differential privacy, we require $e^{\lambda(z)} \geq g(v)_z / g(S_{z,z^+}(v))_z$. Let $v = (z, v_{-i})$ and $\hat{v} = (z^-, v_{-i})$. Note that $r_z(v) = r_z(\hat{v}) + 1$ and $r_{z^-}(v) = r_{z^-}(\hat{v})$. Fix a switch S_{z,z^+} and a rank ρ and consider some environment v_{-i} such that $r_z(z, v_{-i}) = r$. The right side of the inequality becomes

$$\frac{\sum_{r=1}^{r_z(z,v_{-i})} t_z(r) - \sum_{r=1}^{r_{z^-}(z,v_{-i})} t_{z^-}(r) + \xi_z}{\sum_{r=1}^{r_z(z^+,v_{-i})} t_z(r) - \sum_{r=1}^{r_{z^-}(z^+,v_{-i})} t_{z^-}(r) + \xi_z}$$
$$= \frac{t_z(\rho) + \sum_{r=1}^{r_z(v)} t_z(r) - \sum_{r=1}^{r_{z^-}(v)} t_{z^-}(r) + \xi_z}{\sum_{r=1}^{r_z(\hat{v})} t_z(r) - \sum_{r=1}^{r_{z^-}(\hat{v})} t_{z^-}(r) + \xi_z}$$

The largest that this fraction can be is for an environment v_{-i} in which no votes equal z, in which case $r_z(\hat{v}) = r_{z^-}(\hat{v}) = \rho - 1$. Plugging this in and simplifying, we satisfy (λ, B) differential privacy if

$$\frac{t_z(\rho) + \sum_{r=1}^{\rho-1} \big(t_z(r) - t_{z^-}(r)\big) + \xi_z}{\sum_{r=1}^{\rho-1} \big(t_z(r) - t_{z^-}(r)\big) + \xi_z} \leq e^{\lambda z}$$

□

VII. Discussion

We began this study with the observation that we could incorporate user's privacy preferences into an information system by having them vote on the privacy parameters they would like to have. We presented elements that such an ϵ voting system would need. We then asked whether it was possible for a system to aggregate user preferences over ϵ in a reasonable way while also respecting privacy. In fact, our results are quite mixed.

For one thing, user preferences may not always be well-behaved. For example, there may be a threshold value at which an adversary becomes convinced enough of a user's sensitive attribute to act on that information. Utility functions may also be single-troughed instead of single-peaked. This could occur in a setting where an individual has some secret they would like to hide, but once it's out, they'd rather everyone else's secret be transparent too. Theorem 1 states that for general utility functions, there is a very restrictive class of feasible mechanisms for choosing ϵ. These are the ones we call randomized dictatorships with phantoms. As a result, we cannot hope to meet, or even approximate, standards such as welfare maximization in the general case.

There are privacy scenarios in which we may expect utility functions to be better-behaved. Theorem 2 states that when utility functions are single-peaked, we can do better than the general case. In fact, we can implement a broader class of mechanisms we call generalized randomized medians.

Generalized randomized medians exhibit well-behaved properties, especially as the number of users n grows. For example, with large n a generalized randomized median can come close to approximating a simple median, or any other order statistic. These are certainly more responsive to user preferences than randomized dictatorships with phantoms. On the other hand, it is still impossible to approximately maximize welfare without quite strong assumptions about the shape of utility functions.

Our model assumes that users are able to perceive and judge the effects of a privacy parameter. Further work is needed to test the limits of this assumption, how privacy parameters like ϵ may be presented to be salient to a general audience, or how to improve the usability of private algorithms more generally. Our hope is that efforts like this could lead to users that are empowered to understand and affect the use of their personal information.

Acknowledgment

This research was supported in part by UC Berkeley's Center for Long-Term Cybersecurity.

References

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

[2] Ú. Erlingsson, V. Pihur, and A. Korolova, "Rappor: Randomized aggregatable privacy-preserving ordinal response," in *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*. ACM, 2014, pp. 1054–1067.

[3] J. Tang, A. Korolova, X. Bai, X. Wang, and X. Wang, "Privacy loss in apple's implementation of differential privacy on macos 10.12," *arXiv preprint arXiv:1709.02753*, 2017.

[4] M. Naldi and G. D'Acquisto, "Differential privacy: An estimation theory-based method for choosing epsilon," *arXiv preprint arXiv:1510.00917*, 2015.

[5] J. Lee and C. Clifton, "How much is enough? choosing ε for differential privacy," in *International Conference on Information Security*. Springer, 2011, pp. 325–340.

[6] J. Hsu, M. Gaboardi, A. Haeberlen, S. Khanna, A. Narayan, B. C. Pierce, and A. Roth, "Differential privacy: An economic method for choosing epsilon," in *Computer Security Foundations Symposium (CSF), 2014 IEEE 27th*. IEEE, 2014, pp. 398–410.

[7] P. M. Schwartz, "Internet privacy and the state," *Conn. L. Rev.*, vol. 32, p. 815, 1999.
[8] A. Gibbard, "Manipulation of voting schemes: a general result," *Econometrica: journal of the Econometric Society*, pp. 587–601, 1973.
[9] M. A. Satterthwaite, "Strategy-proofness and arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions," *Journal of economic theory*, vol. 10, no. 2, pp. 187–217, 1975.
[10] A. Gibbard, "Manipulation of schemes that mix voting with chance," *Econometrica: Journal of the Econometric Society*, pp. 665–681, 1977.
[11] ——, "Straightforwardness of game forms with lotteries as outcomes," *Econometrica: Journal of the Econometric Society*, pp. 595–614, 1978.
[12] F. McSherry and K. Talwar, "Mechanism design via differential privacy," in *Foundations of Computer Science, 2007. FOCS'07. 48th Annual IEEE Symposium on*. IEEE, 2007, pp. 94–103.
[13] M. M. Pai and A. Roth, "Privacy and mechanism design," *ACM SIGecom Exchanges*, vol. 12, no. 1, pp. 8–29, 2013.
[14] C. Dwork, A. Roth *et al.*, "The algorithmic foundations of differential privacy," *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014.
[15] Z. Huang and S. Kannan, "The exponential mechanism for social welfare: Private, truthful, and nearly optimal," in *Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium on*. IEEE, 2012, pp. 140–149.
[16] 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*. ACM, 2012, pp. 203–213.
[17] D. Xiao, "Is privacy compatible with truthfulness?" in *Proceedings of the 4th conference on Innovations in Theoretical Computer Science*. ACM, 2013, pp. 67–86.
[18] H. Moulin, "On strategy-proofness and single peakedness," *Public Choice*, vol. 35, no. 4, pp. 437–455, 1980.
[19] Y. Chen, S. Chong, I. A. Kash, T. Moran, and S. Vadhan, "Truthful mechanisms for agents that value privacy," *ACM Transactions on Economics and Computation (TEAC)*, vol. 4, no. 3, p. 13, 2016.
[20] C. Dwork, G. N. Rothblum, and S. Vadhan, "Boosting and differential privacy," in *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*. IEEE, 2010, pp. 51–60.

APPENDIX A
PROOF OF ($\implies$) DIRECTION OF THEOREM 1

Step 1: Show that any switch from alternative x to y only alters the probability that g assigns to x and y.

Take any environment v_{-i} and suppose player i switches their vote from $v_i = x$ to y. We claim that $A(S_{xy}, v)_z = 0$ for all $z \in B - \{x, y\}$. By way of contradiction, suppose that there is some $z \in B - \{x, y\}$ such that $A(S_{x,y}, v)_z = 0$. Consider the utility function $u_i(\epsilon) = -\mathbb{I}(\epsilon = z)$. Since both x and y maximize u_i, Lemma 3 implies $u_i \cdot A(S_{x,y}, v) = 0$. But then

$$0 = u_i \cdot A(S_{x,y}, v) = u_i(z) A(S_{x,y}, v)_z \neq 0$$

producing a contradiction. Thus, g must only move probability between x and y.

Step 2: Show that the mass moved by a switch from x to y is the same as the mass moved by a switch from y to x.

By Step 1, a switch between x and y only transfers mass between x and y. So $A(S_{x,y}, v) = m_{x,y}(v)(\delta^y - \delta^x)$ for some constant $m_{x,y}(v) \in \mathbb{R}$. We will refer to $m_{x,y}(v)$ as the mass moved from x to y in under v. Since g is a probability distribution, the gain in probability mass for y from the switch must be offset by a loss in mass for x, and vice-versa. Thus,

$$\begin{aligned} g(S_{x,y}(v))_y - g(v)_y &= -\big(g(S_{x,y}(v))_x - g(v)_x\big) \\ &= g(v)_x - g(S_{x,y}(v))_x \\ &= g(x, v_{-i})_x - g(y, v_{-i})_x \\ &= g(S_{y,x}(v))_x - g(v)_x \end{aligned}$$

yielding $A(S_{x,y}, v)_y = A(S_{y,x}, v)_x$. So $m_{x,y}(v)(\delta^y - \delta^x)_y = m_{y,x}(v)(\delta^x - \delta^y)_x$, implying $m_{x,y}(v) = m_{y,x}(v)$.

Step 3: Deduce that $g(v)_y$ is a function of $n_y(v)$.

Consider $v, w \in B^n$, such that that $n_y(v) = n_y(w)$. Denote $n_y(w) = l$. By anonymity, we can assume without loss of generality that the first $n - l$ components of both v and w are non-y, and the last l components are y.

Utilizing Step 1, we can construct a sequence of $n - l$ switches that transforms v into w while preserving the probability mass at alternative y. In particular, consider $(S_{v_1,w_1}, ..., S_{v_{n-l},w_{n-l}})$. Define T^0 as the identity map, and $T^j = S_{v_j,w_j} \circ ... \circ S_{v_1,w_1}$ for all $j \in \{1, ..., n - l\}$.

Note that each switch S_{v_j,w_j} is a switch between two votes that are not alternative y. As such, Step 1 implies that $A(S_{v_j,w_j}, T^{j-1}(v))_y = 0$. By the definition of effect, it follows that $g(T^j(v))_y = g(T^{j-1}(v))_y$.

By construction, this process terminates with $T^{n-l}(v) = w$. Furthermore, $T^0(v) = v$. Hence,

$$g(v)_y = g(T^0(v))_y = ... = g(T^{n-l}(v))_y = g(w)_y$$

Thus, $g(v)_y$ produces the same output for all inputs with the same number of votes for y. So $g(v)_y$ is some function of the number of votes. Call this function $f_y(n_y(v))$.

Step 4: Show that the mass moved from a switch between x and y is the same regardless of vote vector.

We want to show that for any $v, w \in B^n$, $m_{x,y}(v) = m_{x,y}(w)$. By Step 2, it is sufficient to show that $A(S_{x,y}, v) = A(S_{x,y}, w)$. To do so, we choose a sequence of switches $(T^1, T^2, ..., T^q)$ such that applying them to v in order results in w. One algorithm for constructing such a sequence is the following. Let T be an empty sequence. Looping over all $i \in [n]$,

- If $v_i = w_i$, do nothing
- Else if $v_i \neq w_i$ and $\{v_i, w_i\} \neq \{x, y\}$, add switch S_{v_i,w_i} to the end sequence T
- Else, choose some $z \in B - \{x, y\}$ and add switches $S_{v_i,z}$ and S_{z,w_i} to the end of sequence T in that order

Denote the resulting sequence $T = (T^1, T^2, ..., T^q)$. Notice that we have constructed the sequence T specifically such that none of the switches are $S_{x,y}$ or $S_{y,x}$. For any $\hat{v} \in B^n$, and any T^i in the above sequence, Lemma 1 implies

$$A(S_{x,y} \circ T^i, \hat{v}) = A(T^i \circ S_{x,y}, \hat{v})$$

Using Lemma 2, we can expand both side to produce

$$A(S_{x,y}, T^i(\hat{v})) + A(T^i, \hat{v}) = A(S_{x,y}, \hat{v}) + A(T^i, S_{x,y}(\hat{v}))$$

By Step 2, note that the first terms on each side are vectors that point in the direction $\delta_y - \delta_x$. The second terms also point in the same direction as each other, but in a direction that is not $\delta_y - \delta_x$ by the way we've constructed the sequence T^i. Denote the first term on the left side as W, the second term on the left as X, the first term on the right side as Y, and the second term on the right as Z. Observe that $WXYZ$ forms a quadrilateral in $\mathbb{R}^k$. The above analysis of directions implies that W and Y are parallel to one another, and X and Z are also parallel to one another. Hence $WXYZ$ is a parallelogram, implying that opposite sides must have the same magnitude. In particular, the lengths of W and Y are equal, yieliding $A(S_{x,y}, \hat{v}) = A(S_{x,y}, T^i(\hat{v}))$. Repeated application of this relationship yields

$$\begin{aligned} A(S_{x,y}, v) &= A(S_{x,y}, T^1(v)) = A(S_{x,y}, T^2 \circ T^1(v)) \\ &= ... = A(S_{x,y}, T^q \circ ... \circ T^1 v) = A(S_{x,y}, w) \end{aligned}$$

Step 5: Show that any switch between two alternatives on the ballot results in the transfer of a unique probability mass.

To establish that the transfer of mass is the same between any switch, consider the event where player i switches votes from x to y. By Step 4, this mass is the same regardless of the vote vector v. As such, when referring to $m_{x,y}(v)$ we can omit v. Denote this value as $m_{x,y}$. So by Step 2,

$$m_{x,y} = A(S_{x,y}, v)_x = A(S_{y,x}, v)_y = g(x, v_{-i})_x - g(y, v_{-i})_x$$

Similarly, when player i deviates from x to z,

$$m_{x,z} = A(S_{x,z}, v)_z = A(S_{z,x}, v)_x = g(x, v_{-i})_x - g(z, v_{-i})_x$$

Subtracting these quantities, $m_{x,y} - m_{x,z} = g(z, v_{-i})_x - g(y, v_{-i})_x$. By Step 3, we've established that $g(v)_x = f_x(n_x(v))$. Since $z \neq x$ and $y \neq x$, it follows that $n_x(z, v_{-i}) = n_x(y, v_{-i})$. So, $g(z, v_{-i})_x = g(y, v_{-i})_x$. Hence $m_{x,z} = m_{x,y}$. Thus, any deviation from a candidate x to any other alternative must transfer the same mass.

By Step 2, it follows that $m_{x,y} = m_{y,x}$ and $m_{x,z} = m_{z,x}$, so $m_{y,x} = m_{z,x}$ as well. Thus, any deviation from an alternative that is not z to the candidate z must transfer the same mass. Combining these two results, it follows that any deviation results in the same mass being moved under any switch. Denote this value as m.

Step 6: Deduce that g_z has the desired functional form.

Fix $v_{-i} \in B^{n-1}$ and suppose player i switches their vote from x to y. Then,

$$\begin{aligned} m &= g(y, v_{-i})_y - g(x, v_{-i})_y \\ &= f_y(n_y(y, v_{-i})) - f_y(n_y(x, v_{-i})) \\ &= f_y(n_y(y, v_{-i})) - f_y(n_y(y, v_{-i}) - 1) \\ &= \frac{f_y(n_y(y, v_{-i})) - f_y(n_y(y, v_{-i}) - 1)}{n_y(y, v_{-i}) - (n_y(y, v_{-i}) - 1)} \end{aligned}$$

which implies that $f_y(n_y(v)) = m n_y(v) + b_y$ for some constant b_y.

Next, we establish that all $b_y \geq 0$. By way of contradiction, suppose there is some $y \in B$ such that $b_y < 0$. Consider the case where all individuals vote for some $x \neq y$; i.e. $v = (x, ..., x)$. Then $g(v)_y = m n_y(v) + b_y = b_y < 0$, contradiction to fact that this is a probability.

Furthermore, we claim that $m > 0$. Since g is truthful, if voter i has type α_i, then any deviation to an alternative $x \in B$ yields $u_i \cdot \big(g(\alpha_i, v_{-i}) - g(x, v_{-i})\big) \geq 0$. Performing the same algebra and deductions as in the ($\Longleftarrow$) direction of the proof above, we see that

$$0 \leq u_i \cdot \big(g(\alpha_i, v_{-i}) - g(x, v_{-i})\big) = m\big(u_i(\alpha_i) - u_i(x)\big)$$

Since voter i has type α_i, $\big(u_i(\alpha_i) - u_i(x)\big) \geq 0$. Thus, $m \geq 0$. Since g is non-degenerate, $m \neq 0$. Thus, $m > 0$. As such, define $\phi_z = m^{-1} b_z$. Then,

$$g(v)_z = m\big(n_z(v) + \phi_z\big)$$

Since $g(v)$ is a probability distribution,

$$1 = \sum_{x \in B} m(n_z(v) + \phi_x) \iff m = \left(n + \sum_{x \in B} \phi_x\right)^{-1}$$

Therefore

$$g(v)_z = \frac{n_z(v) + \phi_z}{n + \sum_{x \in B} \phi_x}$$

Step 7: Show that each $\phi_x \geq \left(e^{\lambda x} - 1\right)^{-1}$.

Since g is (λ, B) differentially private, for any vote vectors (y, v_{-i}) and (z, v_{-i}) it follows that

$$\frac{g(v_i = y, v_{-i})_x}{g(v_i = z, v_{-i})_x} = \frac{n_x(v_{-i}) + \mathbb{I}(y = x) + \phi_x}{n_x(v_{-i}) + \mathbb{I}(z = x) + \phi_x} \leq e^{\lambda x}$$

Examining this middle ratio, we see there are four cases:

$$\frac{g(v_i = y, v_{-i})_x}{g(v_i = z, v_{-i})_x} = \begin{cases} 1 & y = x \text{ and } z = x \\ \frac{n_x(v_{-i}) + 1 + \phi_x}{n_x(v_{-i}) + \phi_x} & y = x \text{ and } z \neq x \\ \frac{n_x(v_{-i}) + \phi_x}{n_x(v_{-i}) + 1 + \phi_x} & y \neq x \text{ and } z = x \\ 1 & y \neq x \text{ and } z \neq x \end{cases}$$

Since g satisfies (λ, B) differential privacy, the constant ϕ_x must be restricted to hold for all four cases. Observe that criteria 1, 3, and 4 are less than or equal to 1 for every v_{-i}, so this bound holds for any $\phi_x \geq 0$. For the second criteria to hold, it must be that for every $v_{-i} \in B^{n-1}$,

$$\frac{n_x(v_{-i}) + 1 + \phi_x}{n_x(v_{-i}) + \phi_x} \leq e^{\lambda x}$$

The left hand side can be rewritten as

$$\frac{\big(n_x(v_{-i}) + \phi_x\big) + 1}{n_x(v_{-i}) + \phi_x} = 1 + \frac{1}{n_x(v_{-i}) + \phi_x}$$

so the above inequality reduces to

$$\phi_x \geq \left(e^{\lambda x} - 1\right)^{-1} - n_x(v_{-i})$$

Note that the bound holds for $n_x(v_i) = 0$, so it automatically holds for all other values. Thus, $\phi_x \geq \left(e^{\lambda x} - 1\right)^{-1}$.

APPENDIX B
PROOF OF ($\implies$) DIRECTION OF THEOREM 2

Step 1: Show that a switch between neighboring ϵ's only affects the probability g assigns to those ϵ's.

Let S_{z,z^+} be a switch between z and z^+. We claim that $A(S_{z,z^+}, v) = m_{z,z^+}(v)(\delta^{z^+} - \delta^z)$ for some $m_{z,z^+}(v) \in \mathbb{R}$. By way of contradiction, suppose not. Then there exists some $d \notin \{z, z^+\}$ such that $A(S_{z,z^+}, v)_d \neq 0$. We first the consider the case that $d < z$.

Let $l = \min\{\epsilon \in B | A(S_{z,z^+}, v)_\epsilon \neq 0\}$. As such, $l \leq d < z$. Define a utility function, $u_i(x) = -\mathbb{I}(x \leq l)$. Since both z and z^+ are at the top of i's preference list, Lemma 3 implies $u_i \cdot A(S_{z,z^+}, v) = 0$. But then,

$$0 = u_i \cdot A(S_{z,z^+}, v) = u_i(l) A(S_{z,z^+}, v)_l \neq 0$$

producing a contradiction. When $d > z^+$, a symmetric argument to the one above holds, establishing the claim.

Step 2: Show that the effect of a neighboring switch is independent of the vote vector, and the mass transferred is a function of the rank of the starting point.

To complete this step, we show that for any $v, w \in B^n$ with $r_z(v) = r_z(w)$, $A(S_{z,z^+}, v) = A(S_{z,z^+}, w)$.

Take $v, w \in B^n$ such that $r_z(v) = r_z(w)$. Denote $r_z(w)$ as r. Since g is anonymous, we can assume without loss of generality that the first r components of both v and w are less than or equal to z, and the remaining $n - r$ components are larger than z. We can construct a sequence of switches $T = (T^1, ..., T^q)$ such that applying them in order transforms v into w. One algorithm for constructing such a sequence is the following. Let T be the empty sequence. Looping over all $i \in [n]$,

- If $v_i = w_i$, do nothing
- Else if $v_i < w_i$, add switches $S_{v_i, v_i^+}, ..., S_{w_i^-, w_i}$ to T in that order
- Else, add switches $S_{v_i, v_i^-}, ..., S_{w_i^+, w_i}$ to T in that order

Denote the resulting sequence T as $(T^1, ..., T^q)$. Note that none of the switches T^i in T are S_{z,z^+} or $S_{z^+,z}$ by construction. For any $\hat{v} \in B^n$, and any T^i in the above sequence, Lemma 1 implies

$$A(S_{z,z^+} \circ T^i, \hat{v}) = A(T^i \circ S_{z,z^+}, \hat{v})$$

Using Lemma 2, and utilizing the same parallelogram style argument present in Theorem 1 Step 4 with Step 1 above, it follows that $A(S_{z,z^+}, \hat{v}) = A(S_{z,z^+}, T^i(\hat{v}))$. Repeated application of this relationship yields

$$\begin{aligned} A(S_{z,z^+}, v) &= A(S_{z,z^+}, T^1(v)) = A(S_{z,z^+}, T^2 \circ T^1(v)) \\ &= ... = A(S_{z,z^+}, T^q \circ ... \circ T^1 v) = A(S_{z,z^+}, w) \end{aligned}$$

Thus, $A(S_{z,z^+}, v)$ produces the same output for all inputs v with the same rank for z. So $A(S_{z,z^+}, v)$ is some function of the rank $r_z(v)$. By Step 1, $A(S_{z,z^+}, v) = m_{z,z^+}(v)(\delta^{z^+} - \delta^z)$. Since the direction is uniquely determined by $(\delta^{z^+} - \delta^z)$, the rank can only affect the mass transfer. Denote $m_{z,z^+}(v) = t_z(r_z(v))$ for some function t_z. While this proves the claim, we will find it more useful to describe $m_{z^+,z}(v)$ instead. For any $v_{-i} \in B^{n-1}$, observe that

$$\begin{aligned} m_{z^+,z}(z^+, v_{-i}) &= -A(S_{z^+,z}, (z^+, v_{-i}))_{z^+} \\ &= A(S_{z,z^+}, (z^+, v_{-i}))_z = m_{z,z^+}(z, v_{-i}) \\ &= t_z(r_z(z, v_{-i})) = t_z(r_z(z^+, v_{-i}) + 1) \end{aligned}$$

Thus, $m_{z^+,z}(v) = t_z(r_z(v)+1)$. Furthermore, the truthfulness of g implies that the range of t_z must be non-negative.

Step 3: Deduce that g has the desired functional form.

Let v_0 be the vector of votes $(\epsilon_k, \epsilon_k, ..., \epsilon_k)$. For a given v, we construct a series of switches, $(T^1, ..., T^q)$, such that each T_i is downward neighboring switch, and $v = T^q \circ ... \circ T^1(v_0)$. To do so, consider the following algorithm. Initialize T to be the empty sequence. Loop over all $i \in [n]$:

- If $v_i = \epsilon_k$, do nothing
- Else, add $S_{\epsilon_k, \epsilon_{k-1}}, ..., S_{v_i^+, v_i}$ in order to the end of T

Note that since each T^i is a neighboring transfer, Step 1 guarantees the probability mass of $g(v_0)$ is only altered at the two components where the switch occurred. By Lemma 2,

$$\begin{aligned} &g(v)_z = g(T^q \circ ... \circ T^1(v_0))_z \\ &= g(v_0)_z + A(T^1, v_0)_z + ... + A(T^q, T^{q-1} \circ ... \circ T^1(v_0))_z \end{aligned}$$

First, we consider $z < \epsilon_k$. By Step 1, the only effects that are nonzero at z are those for switches $S_{z^+,z}$ and S_{z,z^-}. Since $r_z(v) = |\{i \in [n] | v_i \leq z\}|$, there exist $r_z(v)$ switches of $S_{z^+,z}$ in T, which transfer probability into $g(v)_z$. Label these $I^1, I^2, ..., I^{r_z(v)}$ in the order in which they occur in T, and let v^i be the vector of votes right before I^i is applied. Notice that before any of the $\{I^i\}$ is applied, all votes are greater than z, so $r_z(v^1) = 0$. By Step 2, the mass moved into z by I^1 is therefore $t_z(0+1) = t_z(1)$. Similarly, right before I^2 is applied, we have $r_z(v^2) = 1$ so $t_z(2)$ mass is moved into z by I^2. This holds for all remaining $I^3, I^2, ..., I^{r_z(v)}$. Hence the total mass moved into $g(v)_z$ by incoming transfers is therefore $\sum_{r=1}^{r_z(v)} t_z(r)$.

By a similar argument, there exist $r_{z^-}(v)$ switches of S_{z,z^-}, which move a total mass of $\sum_{r=1}^{r_{z^-}(v)} t_{z^-}(r)$ away from $g(v)_z$. Defining $\xi_z = g(v_0)_z$, we finally have,

$$g(v)_z = \sum_{r=1}^{r_z(v)} t_z(r) - \sum_{r=1}^{r_{z^-}(v)} t_{z^-}(r) + \xi_z$$

For ϵ_k, we arbitrarily choose $t_{\epsilon_k} : \{1, ..., n\} \to [0, 1]$ and ξ_{ϵ_k} such that $\sum_{r=1}^{n} t_{\epsilon_k}(r) + \xi_{\epsilon_k} = g(v_0)_{\epsilon_k}$. There are no switches in $\{T^i\}$ that transfer mass into k, but there are $r_{\epsilon_{k-1}}(v)$ that transfer mass out. By a similar argument to above, the total mass transferred out is $\sum_{r=1}^{r_{\epsilon_{k-1}}(v)} t_{\epsilon_{k-1}}(r)$, which implies that $g(v)_{\epsilon_k}$ has the correct form.