Approximately Optimal Mechanism Design via Differential Privacy[*]

Kobbi Nissim[†] Rann Smorodinsky[‡] Moshe Tennenholtz[§]

October 16, 2010
Preliminary Draft

Abstract

In this paper we study the implementation challenge in an abstract common values model and an arbitrary social welfare function. We design a mechanism that allows for approximate optimal implementation in ex-post Nash equilibrium. If, furthermore, values are private then the same mechanism is strategy proof. We cast our results onto two specific models: pricing and facility location. The mechanism we design is optimal up to an additive factor of the order of magnitude of the square root of the optimum and involves no utility transfer.

Underlying our mechanism is a lottery between two auxiliary mechanisms - with high probability we actuate a random mechanism that reduces players influence on the choice of the social alternative, while choosing the optimal outcome with high probability. With the complementary probability we actuate a mechanism that is possibly sub-optimal but is incentive compatible. The joint mechanism inherits the desired properties from both.

[*] We thank Amos Fiat and Haim Kaplan for discussions at an early stage of this research. We thank Frank McSherry and Kunal Talwar for helping to clarify issues related to the constructions in [19].

[†] Microsoft Audience Intelligence, Israel, and Department of Computer Science, Ben-Gurion University. Research partly supported by the Israel Science Foundation (grant No. 860/06). kobbi@cs.bgu.ac.il.

[‡] Faculty of Industrial Engineering and Management, Technion –Israel Institute of Technology, Haifa 32000, Israel. This work was supported by Technion VPR grants and the Bernard M. Gordon Center for Systems Engineering at the Technion. rann@ie.technion.ac.il.

[§] Microsoft Israel R&D Center and the Faculty of Industrial Engineering and Management, Technion–Israel Institute of Technology, Haifa 32000, Israel. moshet@microsoft.com.

1 Introduction

Mechanism design (see Mas-Colell, Whinston and Green [16]) deals with the implementation of desired outcomes in a multi-agent system. The outcome of a mechanism may be a price for a good, an allocation of goods to the agents, the decision on a provision of a public good, locating public facilities, etc. The quality of the outcome is measured by some social welfare function. In many instances the literature is concerned with the sum of the agents' valuations for an outcome, but it can take many other forms, such as the revenue of a seller in an auction setting, the social inequality in a market setting and more.

The holy grail of the mechanism design challenge is to design mechanisms which exhibit dominant strategies for the players, and furthermore, once players play their dominant strategies the outcome of the mechanism coincides with maximizing the social welfare function. Without loss of generality we can replace this with the challenge of designing truthful optimal mechanisms, namely where being truthful is dominant, and truthfulness leads to optimality.

As it turns out, such powerful mechanisms do not exist in general. The famous Gibbard-Satterthwaite theorem (Gibbard [12] and Satterthwaite [28]) tells us that for non-restricted settings any non-trivial truthful mechanism is dictatorial. However, if we restrict attention to the social welfare function that is simply the sum of the agents' valuations, then this problem can be overcome by introducing monetary payments. Indeed, in such cases the celebrated Vickrey-Clarke-Groves mechanisms, discovered by Vickrey [33] and generalized by Clarke [6] and Groves [14], guarantee that being truthful is a dominant strategy and the outcome is efficient. Unfortunately, Roberts [25] showed that a similar mechanism cannot be obtained for other social welfare functions. This cul-de-sac induced researchers to 'lower the bar' for mechanism design. One possibility for lowering the bar is to replace the solution concept with a weaker one and a large body of literature on Bayes-Nash implementation has developed (the reader is referred to Mas-Colell et al [16] for further reading) . Another direction is that of approximate implementation where the quest replaces accurate implementation with approximate implementation, while keeping the approximation inaccuracy as low as possible.

The latter research agenda received much attention in recent years in the context of *algorithmic* mechanism design. Interestingly, approximate implementation was initially motivated by the challenge of designing mechanisms that are computationally tractable (In fact, even the VCG mechanism is not computationally tractable). It turns out that approximate implementation can lead to positive results. In fact, such positive results have been recently provided for social welfare functions other than the sum of agents' valuations and in settings where no money is involved (e.g., Procaccia and Tennenholtz [23]).

The mechanism design literature has characterized functions that are truthfully implemented without payments, and studied domains in which non-dictatorial functions can be implemented (some examples are Moulin [20] and Schummer and Vohra [30, 31]). However, no *general* techniques are known for designing mechanisms that are approximately optimal. Consider the facility location problem, as an example, where the social planner needs to locate some facilities, based on agents' reports of their own location. This problem has received extensive attention recently, yet small changes in the model result in different techniques which seem tightly tailored to the specific model assumptions (see Alon et al. [4], Procaccia and Tennenholtz [23] and Wang et al. [34]). Furthermore, the approximation accuracy in many of these models leaves much to be desired.

1.1 Our Contribution

We introduce an abstract mechanism design model where agents have common values and provide a generic technique for approximate implementation of an arbitrary social welfare function. More precisely, we bound the worst case difference between the optimal outcome ('first best') and the expected outcome of our generic mechanism by $O(\sqrt{\frac{\ln n}{n}})$, where n the population size. In addition, our generic construction does not involve utility transfer and money.

Our construction combines two very different random mechanisms:

- With high probability we deploy a mechanism that chooses social alternatives with a probability that is proportional to (the exponent of) the outcome of the social welfare function, assuming players are truthful. This mechanism exhibits two important properties. First, agents have small influence on the outcome of the mechanism and consequently have little influence on their own utility. As a result all strategies, including truthfulness, are ϵ-dominant. Second, under the assumption that players are truthful, alternatives which are nearly optimal are most likely to be chosen. The concrete construction we use follows the Exponential Mechanism presented by McSherry and Talwar [19].
- With vanishing probability we deploy a mechanism which is designed with the goal of eliciting agents' private information, while ignoring the social welfare function.

The combined mechanism inherits the truthfulness property of the second component and the accuracy of the first component, resulting in an excellent approximation of the optimal outcome in dominant strategies.[1]

Our technique is developed for an abstract setting where both the agents' type space as well as the set of social alternatives are discrete. In more concrete settings, however, our techniques extend to continuous models. In particular, whenever the set of types and alternatives permits a discrete and 'dense' subset. We demonstrate our results and the aforementioned extension in two specific settings: (1) Facility location problems, where the social planner is tasked with the optimal location of K facilities in the most efficient way. In this setting we focus on minimizing

[1] In fact, as our model considers discrete type spaces, the results are obtained for strictly dominant strategies.

the social cost which is the sum of agents' distances from the nearest facility. (2) The digital goods pricing model, where a monopolist needs to determine the price for a digital good (goods with zero marginal cost for production) in order to maximize revenue.

Another contribution of our work is an extension of the classical social choice model. In the classical model agents' utilities are expressed as a function of the private information and a social alternative. This modeling choice abstracts away the issue of how agents exploit the social choice made. A novelty of our model is that it considers an additional stage where agents take an action following the choice made by society, out of a given set of actions (hereinafter 'reaction'). We motivate this extension to the standard model with the following examples:

Facility Location: Consider a location problem, where the social planner is faced with the challenge of locating K similar facilities (e.g., schools) based on where the individuals are located. The K locations are determined based on the individuals' announced locations, and, in turn, agents choose which of the K facilities to access (this is their reaction).

Monopolist pricing: A monopolist wants to maximize its revenue based on the demand curve of the potential buyers. Buyers' demand announcement is followed by a price set by the monopoly (the social alternative). In return agents choose whether to buy the good at the set price or not (the agents' reaction).

Centralized Exchange Economy: Consider a population of n agents in a market, where each agent is endowed with some initial bundle (supply) and a demand function for the goods. The social planner determines a price vector (the social alternative), based on the announced demand and supply. Once prices are announced agents can buy and sell according to the price vector (their reaction).

Public Good: As part of its welfare policy the government plans to offer a portfolio of retraining programs for career changes. Based on agents' (announced) preferences a limited portfolio is offered. Once this portfolio is determined agents can enroll into one program only (their 'reaction'). In an abstract sense this example is similar to the facility location problem.

Network Design: As a final example consider the problem of designing a communication or transportation network, based on agents' (privately known) needs. Once a network is built each agent chooses which path in the network to use (its reaction).

The aforementioned examples demonstrate the prevalence of 'reactions' in a typical design problem. Note that if we assume that the set of reactions is a singleton then we are back to the classical model. An example where the set of reactions is a singleton is that of an election. Once a candidate is elected agents utilities are determined (one could argue that even in this example, agents can adapt behavior to the elected candidate). In our construction we allow for the mechanism to restrict the reaction set (we refer to this as 'imposition').

1.2 Related Work

Approximate Efficiency in Large Populations. The basic driving force underlying our construction is ensuring that each agent has a vanishing influence on the outcome of the mechanism as the population grows. In the limit, if players are non-influential, then they might as well be truthful. This idea is not new and has been used by various authors to provide mechanisms that approximate efficient outcomes when the population of players is large. Some examples of work that hinge on a similar principle for large, yet finite populations, are Swinkels [32] who studies auctions, Satterthwaite and Williams [29] and Rustichini, Satterthwaite and Williams [27] who study double auctions, and Al-Najjar and Smorodinsky [3] who study an exchange market. The same principle is even more enhanced in models with a continuum of players, where each agent has no influence on the joint outcome (e.g., Roberts and Postlewaite [26] who study an exchange economy). The mechanisms provided in these papers are designed for accomplishing efficiency, and provide no value for alternative social welfare functions. In contrat, our results hold a for a wide range of social welfare functions and are generic in nature.

Interestingly, a similar argument has also been instrumental to show inefficiency in large population models. Rob [24] uses lack of influence to model the 'tragedy of the commons' and Mailath and Postlewaite [15] use similar arguments to demonstrate 'free-riding, in the context of the voluntary supply of a public good, which eventually leads to inefficiency.

Differential Privacy and Influence. Consider a (possibly random) function that maps a vector of private inputs into an arbitrary domain. The 'influence' of a player is a measure of how much her input can alter the outcome. In recent years this notion has been researched in two communities: Economics (demonstrated above) and Computer Science.

In Computer Science, the cryptography community has been formalizing a discussion on privacy. The notion of *differential privacy*, introduced in Dwork, McSherry, Nissim and Smith [9] and Dwork [7], captures the 'influence' of a single agent on the result of a computation. More accurately, differential privacy stipulates that the influence of any contributor to the computation is bounded in a very strict sense: any change in the input contributed by an individual translates to at most a near-one multiplicative factor in the probability distribution over the set of outcomes. The scope of computations that can be computed in a differentially private manner has grown significantly since the introduction of the concept (the reader is referred to Dwork [8] for a recent survey). In this strand of the literature computations (equivalently mechanisms) that preserve differential privacy are referred to as ϵ-*differentially private* computations.

McSherry and Talwar [19] establish an inspiring connection between differential privacy and mechanism design. They observe that participants (players) that contribute private information to ϵ-differentially private computations have limited influence on the outcome of the computation, and hence have a limited incentive to lie, even if their utility is derived from the joint outcome.

Consequently, in mechanisms that are ϵ-differentially private truth-telling is approximately dominant, regardless of the agent utility functions. McSherry and Talwar introduce the exponential mechanism as a generic ϵ-differentially private mechanism. In addition, they show that whenever agents are truthful the exponential mechanism chooses a social alternative which almost optimizes the the social welfare function. They go on and demonstrate the power of this mechanism in the context of Unlimited Supply Auctions, Attribute Auctions, and Constrained pricing.

The contribution of McSherry and Talwar, although inspiring, leaves much to be desired in terms of mechanism design: (1) Truth telling is ϵ-dominant for the exponential mechanism. On the one hand, lower values of ϵ imply higher compatibility with incentives. On the other hand, lower values deteriorate the approximation results. What is the optimal choice for ϵ? How can these countervailing forces be reconciled? It turns out that the McSherry and Talwar model and results do not provide a framework for analyzing this. (2) McSherry and Talwar claim that truth telling is *approximately* dominant. In fact, a closer look at their work reveals that **all** strategies are approximately dominant, which suggests that truth telling has no intrinsic advantage over any other strategy in their mechanism. (3) In fact, one can demonstrate that misreporting one's private information can actually dominate other strategies, truth-telling included. To make things worse, such dominant strategies may lead to inferior results for the social planner. This is demonstrated in an example provided in appendix A in the context of monopoly pricing.

Abstract formalizations of influence have also been studied in the economic literature. Fudenberg Levine and Pesendorfer [11] and Al-Najar and Smorodinsky [2], for example, define an agents' influence on a mechanism and provide on average influence and on the number of influential players. McLean and Postlewaite [17, 18] introduce the notion of informational smallness, formalizing settings where one player's information is insignificant with respect to the aggregated information.

Facility Location. One of the concrete examples we investigate is the optimal location of facilities. The facility location problem has already been tackled in the context of approximate mechanism design without money, and turned out to lead to interesting challenges. While the single facility location problem exhibits preferences that are single-peaked and can be solved optimally by selecting the median declaration, the 2-facility problem turns out to be non-trivial. Most recently Wang et al [34] introduce a randomized 4-(multiplicative) approximation truthful mechanism for the 2 facility location problem. The techniques introduced here provide much better approximations - in particular we provide an additive $\tilde{O}(n^{-1/3})$ approximation to the average optimal distance between the agents and the facilities.[2]

Following our formalization of the reaction set and of imposition, Fotakis and Tzamos [10] exploit this even further for facility location. The authors use this idea in the context of previously known mechanisms to improve implementation accuracy. The provide constant multiplicative approximation or logarithmic multiplicative approximation, albeit with fully imposing mechanisms.

[2] The notation $\tilde{O}(n^{-1})$ is used to denote convergence to zero at a rate $\frac{\ln(n)}{n}$.

Non discriminatory Pricing of Digital Goods. Another concrete setting where we demonstrate our generic results is a pricing application, where a monopolist sets a single price for goods with zero marginal costs ("digital goods") in order to maximize revenues. Pricing mechanisms for the private values case have been studied by Goldberg et al [13] and Balcan et al [5]. Balcan et al [5] demonstrate a mechanism that is $O(\frac{1}{\sqrt{n}})$-approximately optimal (where n is the population size), compared with our mechanism, which is inferior, and provides a $\tilde{O}(\frac{1}{n^{1/3}})$-approximation. Whereas the mechanism of Balcan et al is ad-hoc, our mechanism is derived from general principles and therefore more robust. In addition, our mechanism extends to some settings with common values.

Virtual implementation. Another strand of the literature that combines approximation into mechanism design is that of 'virtual implementation' due to Abreu and Matsushima [1]. A function is virtually implementable if for any $\epsilon > 0$ there exists a mechanism that ϵ-approximates the function in equilibrium. Whereas, we focus on settings with asymmetric information among the agents the literature on virtual implementation mainly looks at settings with on symmetric information, where only the social planner is less informed.

2 Model

2.1 The Environment

Let N denote a set of n agents, S denotes a *finite* set of social alternatives and T_i, $i = 1, \ldots, n$, is a finite type space for agent i. We denote by $T = \times_{i=1}^{n} T_i$ the set of type tuples and by $T_{-i} = \times_{j \neq i} T_j$. Agent i's type, $t_i \in T_i$, is her private information and is known only to her. Let R_i be the set of reactions available to i. Typically, once a social alternative, $s \in S$, is determined agents choose a reaction $r_i \in R_i$. The utility of an agent i is therefore a function of the vector of types, the chosen social alternative and the chosen reaction. Formally, $u_i : T \times S \times R_i \to [0, 1]$.[3] A tuple (T, S, R, u), where $R = \times_{i=1}^{n} R_i$ and $u = (u_1, \ldots, u_n)$, is called an *environment*.

We say that an agent has private reactions if her optimal reaction of i depends only only on her type and the social alternative. Formally, agent i has *private reactions* if $\mathit{argmax}_{r_i \in R_i} u_i((t_i, t_{-i}), s, r_i) = \mathit{argmax}_{r_i \in R_i} u_i((t_i, t'_{-i}), s, r_i)$, for all s, i, t_i, t_{-i} and t'_{-i}. We will use $r_i(t_i, s)$ to denote an arbitrary optimal reaction (i.e., $r_i(t_i, s)$ is an arbitrary function which image is in the set $\mathit{argmax}_{r_i \in R_i} u_i(t_i, s, r_i)$). We say that an agent has *private values* if has private reactions and furthermore $u_i((t_i, t_{-i}), s, r_i) = u_i((t_i, t'_{-i}), s, r_i)$ for all s, i, t_i, t_{-i} and t'_{-i}.

We say that an environment is *non-trivial* if for any pair of types there exists a social alternative for which the optimal reactions are distinct. Formally, $\forall i$, $t_i \neq \hat{t}_i \in T_i$ and t_{-i} there exists $s \in S$,

[3] Utilities are assumed to be bounded in the unit interval. This is without loss of generality, as long as there is some uniform bound on the utility.

denoted $s(t_i, \hat{t}_i, t_{-i})$, such that *argmax*$_{r_i \in R_i} u_i((t_i, t_{-i}), s, r_i) \cap$ *argmax*$_{r_i \in R_i} u_i((\hat{t}_i, t_{-i}), s, r_i) = \emptyset$. We say that $s(t_i, \hat{t}_i, t_{-i})$ *separates* between t_i and $\hat{t}_i$ at t_{-i}. A set of social alternatives, $\tilde{S} \subset S$ is called *separating* if for any i and $t_i \neq \hat{t}_i$ and t_{-i}, there exists some $s(t_i, \hat{t}_i, t_{-i}) \in \tilde{S}$ that separates between t_i and $\hat{t}_i$ at t_{-i}.

2.2 The Social Welfare Function

A social planner, not knowing the vector of types, wants to maximize an arbitrary *social welfare function*, $F : T \times S \to [0, 1]$.[4] We focus our attention on a class of functions for which individual agents have a diminishing impact, as the population size grows:

Definition 1 (Sensitivity) *The social welfare function $F : T \times S \to [0, 1]$ is* d-sensitive *if $\forall i, t_i \neq \hat{t}_i, t_{-i}$ and $s \in S$, $|F((t_i, t_{-i}), s) - F((\hat{t}_i, t_{-i}), s)| \leq \frac{d}{n}$, where n is the population size.*

Note that this definition refers to unilateral changes in announcements, while keeping the social alternative fixed. In particular d-sensitivity does not exclude the possibility of a radical change in the optimal social alternative as a result of unilateral deviations, which, in turn, can radically change the utility of the player. Thus, this definition is mute in the context of the influence of an agent on her own utility.

One commonly used social welfare function which is 1-sensitive is the average utility, $F(t, s) = \frac{\sum_i u_i(t,s)}{n}$. Note that a d-sensitive function eliminates situations where any single agent has an overwhelming impact on the value of the social welfare function, for a fixed social alternative s. In fact, if a social welfare function is not d-sensitive, for any d, then in a large population this function could be susceptible to minor faults in the system (e.g., noisy communication channels). Most examples of social welfare function studied in the literature are d-sensitive.[5]

2.3 Mechanisms

Denote by $\mathcal{R}_i = 2^{R_i} \setminus \{\emptyset\}$ the set of all subsets of R_i, except for the empty set, and let $\mathcal{R} = \times_i \mathcal{R}_i$.

A (direct) mechanism randomly chooses, for any vector of inputs t a social alternative, and for each agent i a subset of available reactions. Formally:

Definition 2 (Mechanism) *A (direct)* mechanism *is a function $M : T \to \Delta(S \times \mathcal{R})$.*

[4] Bounding F within the unit interval is without loss of generality and can be replaced with any finite interval.

[5] A fabricated example of a function that is not d-sensitive is a parity based function such as setting $F = 1$ if the number of some type is odd, and $F = 0$ when it is even.

In addition to publicly announcing the social alternative the mechanism may reveal some information to agents about opponents' announcements. In this paper we shall assume that the mechanism discloses the true vector of announced types to all agents.[6]

We denote by $M_S(t)$ the marginal distribution of $M(t)$ on S and by $M_i(t)$ the marginal distribution on $\mathcal{R}_i$. We say that the mechanism M is *non-imposing* if $M_i(t)(R_i) = 1$. That is, the probability assigned to the grand set of reactions is one, for all i and $t \in T$. Put differently, the mechanism never restricts the set of available reactions. M is *ϵ-imposing* if $M_i(t)(R_i) \geq 1 - \epsilon$ for all i and $t \in T$. In words, with probability exceeding $1 - \epsilon$ the mechanism poses no restrictions.

2.4 Strategies and Solution Concepts

A mechanism induces the following game with incomplete information. In the first phase agents announce their types simultaneously to the mechanism. Then the mechanism chooses a social alternative and a subset of reactions for each agent. In the second stage of the game each agent, knowing the the strategy tuple of all agents, the vector of announced types and the social alternative, must choose a reaction. As our analysis focuses on rational behavior we assume agents choose some reaction in *argmax*$_{r_i \in R_i} u_i((t_i, W_{-i}(t_{-i})), s, r_i)$. Therefore we suppress the reference to the choice of reaction in our notation of a strategy.

A (pure) strategy for i is a function $W : T_i \to T_i$. Note that with this shorthand the action of agents in the second stage, in case of opponents' deviation, is derived from the original strategy profile and ignores the deviation.

Given a vector of types, t, and a strategy tuple W, the mechanism M induces a probability distribution, $M(W(t))$ over the set of social alternatives and reaction tuples. The expected utility of i, at a vector of types t, is $E_{M(W(t))} u_i(t, s, r_i)$.

A strategy W_i is *dominant* for the mechanism M if for any i, any vector of types $t \in T$, any alternative strategy $\hat{W}_i$ of i and any strategy profile $\bar{W}_{-i}$ of i's opponents, the following holds: $E_{M((W_i(t_i), \bar{W}_{-i}(t_{-i})))} u_i(t, s, r_i) \geq E_{M((\hat{W}_i(t_i), \bar{W}_{-i}(t_{-i})))} u_i(t, s, r_i)$. In words, W_i is a strategy that maximizes the expected payoff of i for any vector of types and any strategy used by her opponents. If for all i the strategy $W_i(t_i) = t_i$ is dominant then M is called *truthful* (or *incentive compatible*).[7]

A strategy W_i is *strictly dominant* if it is dominant and furthermore whenever $W(t_i) \neq \hat{W}(t_i)$ then a strong inequality holds. If $W_i(t_i) = t_i$ is strictly dominant then for all i then M is *strictly truthful.*

Finally, a strategy tuple W is an *ex-post Nash Equilibrium* if for all i and $t \in T$ and for any strategy $\hat{W}_i$ of player i, $E_{M(W(t))} u_i(t, s, r_i) \geq E_{M((\hat{W}_i(t_i), W_{-i}(t_{-i})))} u_i(t, s, r_i)$. If $\{W_i(t_i) = t_i\}_{i=1}^n$ is an ex-post Nash equilibrium then M is *ex-post Nash truthful.*

[6] If all agents have private reactions then the public announcement of the agents' announcements is redundant.

[7] Note we do not require a strong inequality to hold on any instance.

2.5 Implementation

Given a vector of types, t, the expected value of the social welfare function, F, at the strategy tuple W is $E_{M(W(t))}[F(t,s)]$.

Definition 3 (β-implementation) *We say that the mechanism M β-*implements F in (strictly) dominant strategies, *for $\beta > 0$, if for any (strictly) dominant strategy tuple, W, for any $t \in T$, $E_{M(W(t))}[F(t,s)] \geq max_{s \in S} F(t,s) - \beta$. A mechanism M β-*implements F in an ex-post Nash equilibrium *if for some ex-post Nash equilibrium strategy tuple, W, for any $t \in T$, $E_{M(W(t))}[F(t,s)] \geq max_{s \in S} F(t,s) - \beta$.*

Main Theorem (informal statement): For any d-sensitive function F and $1 > \beta > 0$ there exists a number n_0 and a mechanism M which β-implements F in an ex-post Nash equilibrium, whenever the population has more than n_0 agents. If, in addition, reactions are private then M β-implements F in strictly dominant strategies.

In addition to a generic mechanism and result we study two specific models and cast our generic results onto those settings. In both models, facility location and pricing, we derive new results.

3 A Framework of Approximate Implementation

In this section we present a general scheme for implementing arbitrary social welfare functions in large societies. The convergence rate we demonstrate is of an order of magnitude of $\sqrt{\frac{\ln(n)}{n}}$. Our scheme involves a lottery between two mechanisms: (1) The Exponential mechanism, a non-imposing mechanism that randomly selects a social alternative in exponential proportion to the value it induces on F; and (2) The Commitment mechanism which is imposing but ignores agents' announcements when (randomly) selecting a social alternative.

3.1 The Exponential Mechanism and Differential Privacy

Consider the following non-imposing mechanism, which we refer to as the Exponential mechanism, originally introduced by McSherry and Talwar [19]:

$$M^{\epsilon}(t)(s) = \frac{e^{n\epsilon F(t,s)}}{\sum_{\bar{s} \in S} e^{n\epsilon F(t,\bar{s})}}.$$

The Exponential mechanism has two notable properties, as we show below. First, it provides ϵ-differential privacy, which is inspired by the literature on privacy. We follow Dwork et al [9] and define:

Definition 4 (ϵ-differential privacy) *A mechanism, M, provides* ϵ-differential privacy *if it is non-imposing and for any $s \in S$, any pair of type vectors $t, \hat{t} \in T$, which differ only on a single coordinate, $M(t)(s) \leq e^{\epsilon} \cdot M(\hat{t})(s)$.*[8]

In words, a mechanism preserves ϵ-differential privacy if for any vector of announcements a unilateral deviation changes the probabilities assigned to any social choice $s \in S$ by a (multiplicative) factor of e^{ϵ}, which approaches 1 as ϵ approaches zero.[9]

Lemma 1 (McSherry and Talwar [19]) *If F is d-sensitive then $M^{\frac{\epsilon}{2d}}(t)$ preserves ϵ-differential privacy*

The proof is simple, and is provided for completeness:

Proof: Let t and $\hat{t}$ be or two type vectors that differ on a single coordinate. Then for any $s \in S$, $F(t,s) - \frac{d}{n} \leq F(\hat{t},s) \leq F(t,s) + \frac{d}{n}$, hence,

$$\frac{M^{\frac{\epsilon}{2d}}(t)(s)}{M^{\frac{\epsilon}{2d}}(\hat{t})(s)} = \frac{\frac{e^{\frac{n\epsilon F(t,s)}{2d}}}{\sum_{\bar{s}\in S} e^{\frac{n\epsilon F(t,\bar{s})}{2d}}}}{\frac{e^{\frac{n\epsilon F(\hat{t},s)}{2d}}}{\sum_{\bar{s}\in S} e^{\frac{n\epsilon F(\hat{t},\bar{s})}{2d}}}} \leq \frac{\frac{e^{\frac{n\epsilon F(t,s)}{2d}}}{\sum_{\bar{s}\in S} e^{\frac{n\epsilon F(t,\bar{s})}{2d}}}}{\frac{e^{\frac{n\epsilon (F(t,s)-\frac{d}{n})}{2d}}}{\sum_{\bar{s}\in S} e^{\frac{n\epsilon (F(t,\bar{s})+\frac{d}{n})}{2d}}}} = e^{\epsilon}.$$

QED

The appeal of mechanisms that provide ϵ-differential privacy is that they induce near indifference among some strategies, in the following sense:

Lemma 2 *If M in non-imposing and provides ϵ-differential privacy, for some $\epsilon < 1$, then for any agent i, any type tuple t, any strategy tuple W and any alternative strategy for i, $\hat{W}_i$ the following holds:*

$$|E_{M(W(t))}[u_i(t,s,r_i)] - E_{M(\hat{W}_i(t_i),W_{-i}(t_{-i}))}[u_i(t,s,r_i)]| < 2\epsilon.$$

To see this note that $(e^{\epsilon} - 1) \leq 2\epsilon$ whenever $\epsilon < 1$ and recall that u_i returns values in $[0,1]$.

McSherry and Talwar [19] note in particular that in the case of private values truthfulness is 2ϵ-dominant, which is an immediate corollary of Lemma 2. They combine this with the following observation to conclude that exponential mechanisms approximately implement F in ϵ- dominant strategies:

[8] For non discrete sets the definition requires that $\frac{M(t)(\hat{S})}{M(\hat{t})(\hat{S})} \leq e^{\epsilon}\ \forall \hat{S} \subset S$.

[9] The motivation underlying this definition of ϵ-differential privacy is that if a single agent's input to a database changes then a query on that database would result in (distributionally) similar results. This, in return, suggests that it is difficult to learn new information about the agent from the query, thus preserving her privacy.

Lemma 3 (McSherry and Talwar [19]) *Let $F : T^n \times S \rightarrow [0,1]$ be an arbitrary d-sensitive social welfare function and $n > \frac{e2d}{\epsilon|S|}$. Then for any t, $E_{M^{\frac{\epsilon}{2d}}(t)}[F(t,s)] \geq \max_s F(t,s) - \frac{4d}{n\epsilon}\ln\left(\frac{n\epsilon|S|}{2d}\right)$.*

In particular, this suggests that $\lim_{n\rightarrow\infty} \frac{4d}{n\epsilon}\ln\left(\frac{n\epsilon|S|}{2d}\right) = 0$ whenever the parameters d, ϵ and $|S|$ are held fixed.[10] Therefore, the exponential mechanism is almost optimal for a large and truthful population. The proof is provided to completeness:

Proof: Let $\delta = \frac{2d}{n\epsilon}\ln\left(\frac{n\epsilon|S|}{2d}\right)$. As $n > \frac{e2d}{\epsilon|S|}$ we conclude that $\ln\left(\frac{n\epsilon|S|}{2d}\right) > e > 0$ and, in particular, $\delta > 0$.

Fix a vector of types, t and denote by $\hat{S} = \{\hat{s} \in S : F(t,\hat{s}) < \max_s F(t,s) - \delta\}$. For any $\hat{s} \in \hat{S}$ the following holds:

$$M^{\frac{\epsilon}{2d}}(t)(\hat{s}) = \frac{e^{\frac{n\epsilon F(t,\hat{s})}{2d}}}{\sum_{s' \in S} e^{\frac{n\epsilon F(t,s')}{2d}}} \leq \frac{e^{\frac{n\epsilon(\max_s F(t,s)-\delta)}{2d}}}{e^{\frac{n\epsilon \max_s F(t,s)}{2d}}} = e^{-\frac{n\epsilon}{2d}\delta}.$$

Therefore, $M^{\frac{\epsilon}{2d}}(t)(\hat{S}) = \sum_{\hat{s}\in\hat{S}} M^{\frac{\epsilon}{2d}}(t)(\hat{s}) \leq |\hat{S}|e^{-\frac{n\epsilon}{2d}\delta} \leq |S|e^{-\frac{n\epsilon}{2d}\delta}$. Which, in turn, implies:

$$E_{M^{\frac{\epsilon}{2d}}(t)}[F(t,s)] \geq (\max_s F(t,s) - \delta)(1 - |S|e^{-\frac{n\epsilon}{2d}\delta}) \geq \max_s F(t,s) - \delta - |S|e^{-\frac{n\epsilon}{2d}\delta}.$$

Substituting for δ we get that

$$E_{M^{\frac{\epsilon}{2d}}(t)}[F(t,s)] \geq \max_s F(t,s) - \frac{2d}{n\epsilon}\ln\left(\frac{n\epsilon|S|}{2d}\right) - \frac{2d}{n\epsilon}.$$

In addition, $n > \frac{e2d}{\epsilon|S|}$ which implies $\ln\left(\frac{n\epsilon|S|}{2d}\right) > \ln(e) = 1$, and hence $\frac{2d}{n\epsilon} \leq \frac{2d}{n\epsilon}\ln\left(\frac{n\epsilon|S|}{2d}\right)$. Plugging this into the previous inequality yields $E_{M^{\frac{\epsilon}{2d}}(t)}[F(t,s)] \geq \max_s F(t,s) - \frac{4d}{n\epsilon}\ln\left(\frac{n\epsilon|S|}{2d}\right)$ as desired.

QED

Remark: There are other mechanisms which exhibit the similar properties to those of the Exponential Mechanism, namely 'almost indifference' and 'approximate optimality'. The literature on differential privacy is rich in techniques for establishing mechanism with such properties. Some examples are the addition of noise calibrated to global sensitivity by Dwork et al. [9], the addition of noise calibrated to smooth sensitivity and the sample and aggregate framework by Nissim et al. [22]. The reader is further referred to the recent survey of Dwork [8].

[10] This limit also approaches zero if $d, \epsilon, |S|$ depend on n, as long as d/ϵ is sublinear in n and $|S|$ is subexponential in n.

3.2 The Commitment Mechanism

We now consider an imposing mechanism that chooses $s \in S$ randomly, while ignoring agents' announcements. Once s is chosen the mechanism restricts the allowable reactions for i to those that are optimal assuming all agents are truthful. Formally, if s is chosen according to the probability distribution P, let M^P denote the following mechanism: $M_S^P(t)(s) = P(s)$ and $M_i^P(t)(r_i(t,s))|s) = 1$. Players do not influence the choice of s in M^P and so they are (weakly) better off being truthful.

We define the *gap* of the environment, $\gamma = g(T, S, A, u)$, as:

$$\gamma = g(T, S, A, u) = \min_{i, t_i \neq b_i, t_{-i}} \max_{s \in S} \left(u_i(t_i, s, r_i(t, s)) - u_i(t_i, s, r_i((b_i, t_{-i}), s)) \right).$$

In words, γ is a lower bound for the loss incurred by misreporting in case of an adversarial choice of $s \in S$. In non-trivial environments $\gamma > 0$. We say the a distribution P is *separating* if there exists a separating set $\tilde{S} \subset S$ such that $P(\tilde{s}) > 0$ for all $\tilde{s} \in \tilde{S}$. In this case we also say that M^P is a separating mechanism. In particular let $\tilde{p} = \min_{s \in \tilde{S}} P(s)$. The following is straightforward:

Lemma 4 *If the environment (T, S, A, u) is non-trivial and P is a separating distribution over S then $\forall b_i \neq t_i, t_{-i}$, $E_{M^P(t_i, t_{-i})}[u_i(t_i, s)] \geq E_{M^P(b_i, t_{-i})}[u_i(t_i, s)] + \tilde{p}\gamma$.*

Proof: For any pair $b_i \neq t_i$ there exists some $s = s(t_i, b_i)$ for which $u_i(t_i, s, r_i(t_i, s)) \geq u_i(t_i, s, r_i(b_i, s)) + \gamma$. P is separating and so $P(s) \geq \tilde{p}$. Therefore, for any i, any $b_i \neq t_i \in T_i$ and for any t_{-i}, $E_{M^P(t_i, t_{-i})}[u_i(t_i, s)] \geq E_{M^P(b_i, t_{-i})}[u_i(t_i, s)] + \tilde{p}\gamma$, as claimed.

QED

So it is optimal to be truthful. More accurately:

Corollary 1 *If the environment (T, S, A, u) is non-trivial and P is a separating distribution over S then*

1. *Truthfulness is an ex-post Nash equilibrium.*
2. *If agent i has private reactions then truthfulness is a strictly dominant strategy.*

Unfortunately, truthfulness is not a dominant strategy whenever an agent does not have private reactions. Recall that an agent's strategy is composed of an announcement of a type followed by a choice of reaction. The latter stage involves a unilateral decision problem which takes into account the vector of strategies. If an agent deviates the other agents still take into the account her original strategy when solving for the optimal reaction. Therefore, in case of a deviation agents may mis-calculate the optimal reaction.

3.3 A Generic and Nearly Optimal Mechanism

Fix a non-trivial environment (T, S, A, u) with a gap γ, separating set $\tilde{S}$, a d-sensitive social welfare function F and a separating commitment mechanism, M^P, with $\tilde{p} = \min_{s \in \tilde{S}} P(s)$.

Set $\bar{M}_q^\epsilon(t) = (1-q)M^{\frac{\epsilon}{2d}}(t) + qM^P(t)$.

Theorem 1 *If $q\tilde{p}\gamma \geq 2\epsilon$ then the mechanism $\bar{M}_q^\epsilon$ is ex-post Nash truthful. Furthermore, if agents have private reactions then $\bar{M}_q^\epsilon$ is strictly truthful.*

Proof: Follows immediately from Lemma 2 (set $W(t_i) = t_i$), Lemma 4 and Corollary 1.

QED

Set the parameters of the mechanism $\bar{M}_q^\epsilon(t)$ as follows:

- $\epsilon = \sqrt{\frac{\tilde{p}\gamma d}{n}} \sqrt{\ln\left(\frac{n\tilde{p}\gamma|S|}{2d}\right)}$
- $q = \frac{2\epsilon}{\tilde{p}\gamma}$

and consider populations of size $n > n_0$, where n_0 is the minimal integer satisfying $n_0 \geq \max\{\frac{8d}{\tilde{p}\gamma}\ln\left(\frac{\tilde{p}\gamma|S|}{2d}\right), \frac{4e^2d}{\tilde{p}\gamma|S|^2}\}$ and $\frac{n_0}{\ln(n_0)} > \frac{8d}{\tilde{p}\gamma}$.

Lemma 5 *If $n > n_0$ then*

1. $q = \frac{2\epsilon}{\tilde{p}\gamma} < 1$
2. $\epsilon < \tilde{p}\gamma$
3. $n > \frac{2ed}{\epsilon|S|}$

Proof: Part (1): $\frac{n}{\ln(n)} > \frac{n_0}{\ln(n_0)} \geq \frac{8d}{\tilde{p}\gamma}$ which implies $n > \frac{8d}{\tilde{p}\gamma}\ln(n)$. In addition, $n > n_0 > \frac{8d}{\tilde{p}\gamma}\ln\left(\frac{\tilde{p}\gamma|S|}{2d}\right)$. Therefore $n > \frac{4d}{\tilde{p}\gamma}\ln\left(\frac{\tilde{p}\gamma|S|}{2d}\right) + \frac{4d}{\tilde{p}\gamma}\ln(n) = \frac{4d}{\tilde{p}\gamma}\ln\left(\frac{\tilde{p}\gamma|S|n}{2d}\right) \implies (\tilde{p}\gamma)^2 > \frac{4\tilde{p}\gamma d}{n}\ln\left(\frac{\tilde{p}\gamma|S|n}{2d}\right)$. Taking the square root and substituting for ϵ on the right hand side yields $\tilde{p}\gamma > 2\epsilon$ and the claim follows.

Part (2) follows directly from part (1)

Part (3): $n > n_0 \geq \frac{4e^2d}{\tilde{p}\gamma|S|^2} \implies \sqrt{n} > \frac{2ed}{\sqrt{\tilde{p}\gamma d}|S|}$. In addition $n > \frac{4e^2d}{\tilde{p}\gamma|S|} > \frac{2de}{\tilde{p}\gamma|S|}$ which implies $1 < \ln\left(\frac{\tilde{p}\gamma|S|n}{2d}\right)$. Combining these two inequalities we get: $\sqrt{n} > \frac{2ed}{\sqrt{\tilde{p}\gamma d}\sqrt{\ln\left(\frac{\tilde{p}\gamma|S|n}{2d}\right)}|S|}$. Multiplying both sides by $\sqrt{n}$ implies $n > \frac{2ed\sqrt{n}}{\sqrt{\tilde{p}\gamma d}\sqrt{\ln\left(\frac{\tilde{p}\gamma|S|n}{2d}\right)}|S|} = \frac{2ed}{\epsilon|S|}$.

QED

Set $\hat{M}(t) = \bar{M}_q^\epsilon(t)$. Our main result is:

Theorem 2 **(Main Theorem)** *The mechanism $\hat{M}(t)$ is ex-post Nash truthful and, in addition, it $6\sqrt{\frac{d}{\tilde{p}\gamma n}}\sqrt{\ln\left(\frac{n\tilde{p}\gamma|S|}{2d}\right)}$-implements F in ex-post Nash equilibrium, for $n > n_0$. If agents have private reactions the mechanism is strictly truthful and $6\sqrt{\frac{d}{\tilde{p}\gamma n}}\sqrt{\ln\left(\frac{n\tilde{p}\gamma|S|}{2d}\right)}$-implements F in strictly dominant strategies.*

Recall that for ex-post Nash implementation we only need to show that one ex-post Nash equilibrium yields the desired outcome.

Proof: Given the choice of parameters ϵ and q then, Theorem 1 guarantees that $\hat{M}(t)$ is ex-post Nash truthful (and truthful whenever reactions are private). Therefore, it is sufficient to show that for any type vector t,

$$E_{\hat{M}(t)}(F(t,s)) \geq \max_s F(t,s) - 6\sqrt{\frac{d}{\tilde{p}\gamma n}}\sqrt{\ln\left(\frac{n\tilde{p}\gamma|S|}{2d}\right)}.$$

Note that as F is positive, $E_{M^P(t)}[F(t,s)] \geq 0$ and so

$$E_{\hat{M}(t)}[F(t,s)] \quad \geq \quad (1-q)E_{M^{\frac{\epsilon}{2d}}(t)}[F(t,s)].$$

By part (3) of Lemma 5 we are guaranteed that the condition on the size of of the population of Lemma 3 holds and so we can apply Lemma 3 to conclude that:

$$E_{\hat{M}(t)}[F(t,s)] \geq (1-q)\left(\max_s F(t,s) - \frac{4d}{n\epsilon}\ln\left(\frac{n\epsilon|S|}{2d}\right)\right).$$

We substitute q with $\frac{2\epsilon}{\tilde{p}\gamma}$ and recall that $\max_s F(t,s) \leq 1$. In addition, part (1) of Lemma 5 asserts that $\frac{2\epsilon}{\tilde{p}\gamma} < 1$. Therefore

$$E_{\hat{M}(t)}[F(t,s)] \geq \max_s F(t,s) - \frac{2\epsilon}{\tilde{p}\gamma} - \frac{4d}{n\epsilon}\ln\left(\frac{n\epsilon|S|}{2d}\right) \geq \max_s F(t,s) - \frac{2\epsilon}{\tilde{p}\gamma} - \frac{4d}{n\epsilon}\ln\left(\frac{n\tilde{p}\gamma|S|}{2d}\right),$$

where the last inequality is based on the fact $\epsilon < \tilde{p}\gamma$, which is guaranteed by part (2) of Lemma 5. Substituting ϵ for $\sqrt{\frac{\tilde{p}\gamma d}{n}}\sqrt{\ln\left(\frac{n\tilde{p}\gamma|S|}{2d}\right)}$ we conclude that

$$E_{\hat{M}(t)}[F(t,s)] \geq \max_s F(t,s) - 2\sqrt{\frac{d}{\tilde{p}\gamma n}}\sqrt{\ln\left(\frac{n\tilde{p}\gamma|S|}{2d}\right)} - 4\sqrt{\frac{d}{\tilde{p}\gamma n}}\sqrt{\ln\left(\frac{n\tilde{p}\gamma|S|}{2d}\right)}$$

and the result follows.

QED

One particular case of interest is the commitment mechanism M^U, where U is the uniform distribution over the set S:

Corollary 2 *Let n_0 be the minimal integer satisfying $n_0 \geq \max\{\frac{8\tilde{d}|S|}{\gamma}\ln\left(\frac{\gamma}{2d}\right), \frac{4e^2 d}{\gamma|S|}\}$ and $\frac{n_0}{\ln(n_0)} > \frac{8d|S|}{\gamma}$. Then the mechanism $\hat{M}^U(t)$ $6\sqrt{\frac{d|S|}{\gamma n}}\sqrt{\ln\left(\frac{n\gamma}{2d}\right)}$-implements F, $\forall n > n_0$.*

Proof: $P = U$ implies that the minimal probability is $\tilde{p} = \frac{1}{|S|}$. Plugging this into Theorem 2 gives the result.

QED

Holding the parameters of the environment $d, \gamma, |S|$ fixed the approximation inaccuracy of our mechanism converges to zero at a rate of $\sqrt{\frac{\ln(n)}{n}}$. However, the dependence on the size of the set of alternatives, S, may be different, as we take different commitment mechanisms. We take advantage of this observation in the analysis of two applications.

4 Pricing Digital Goods

A monopolist producing digital goods, for which the marginal cost of production is zero, faces a set of indistinguishable buyers. Each buyer has a unit demand with a valuation in the unit interval. Each agent is a member of a (small) cohort of agents and all cohort members have the same valuation. The private information of each agent is a signal that is correlated with this valuation. The monopolist wants to set a uniform price in order to maximize her average revenue per user.

Assume there are $N \cdot D$ agents, with agents labeled (n, d) (the d^{th} agent in the n^{th} cohort). The valuation of all agents is the n^{th} cohort is the same and is denoted V^n. Agent (n, d) receives a signal (her type) $X_d^n \in \mathbb{R}$ that is positively correlated with V^n. Given the vector $X^n = \{X_d^n\}_{d=1}^D$, of all the signals of agents in the n^{th} cohort, we denote the expected value of the good to agent (n, d) by $E(V^n|X^n)$ (note that although we suppress the notation of the prior distribution over the valuations and the signals, the conditional expectation refers to them). We assume that each agent's signal is informative in the sense that $E(V^n|X^n) > E(V^n|\hat{X}^n)$ whenever $X^n > \hat{X}^n$ (in each coordinate a weak inequality holds and for at least one of the coordinates a strong inequality holds). That is, whenever an individual's signal increases the expected valuation increases.

Let $R_i = \{$'Buy', 'Not buy'$\}$, be i's set of reactions.

The (expected) utility of (n, d), given the vector of signals $X = \{X^n\}_{n=1}^N = \{\{X_d^n\}_{d=1}^D\}_{n=1}^N$, and the price p, is

$$u_i(X, p, r_i) = \begin{cases} E(V^n|X^n) - p & \text{if } r_i = \text{'Buy'}, \\ 0 & \text{if } r_i = \text{'Not buy'}. \end{cases}$$

We assume that all valuations are restricted to the unit interval, prices are restricted to some finite grid $S = S_M = \{0, \frac{1}{m}, \frac{2}{M}, \ldots, 1\}$ and X^n_d take on only finitely many values. We assume the price grid is fine enough so that for any two vectors $X^n > \hat{X}^n$ there exists some price $p \in S$ such that $E(V^n|X^n) > p + \frac{2}{m} > p > E(V^n|\hat{X}^n)$. Therefore any vectors of announcements there exists a maximal price for which optimal reaction is Buy. For that price, if an agent announces a lower value then the best reaction would be Not Buy, which will yield a loss of $\frac{1}{m}$ at least. Similarly, there exists the lowest price for which the optimal reaction is Not Buy. Announcing a higher value will result in the optimal reaction being Buy, which yields a loss of $\frac{1}{m}$ at least. We conclude that the gap is $\gamma = \frac{1}{m}$.

The monopolist wants to maximize the average revenue per buyer . Denote by $F(t,p) = \frac{p}{n} \cdot |\{(n,d) : E(V^n|X^n) > p\}|$ and note that a unilateral change in the type of one agent may change at most the buying behavior of d members in her cohort. Therefore, F is d-sensitive.

In this application the gap is equal $\frac{1}{m}$. To see this consider the case where an agent announces some $b^n_d \geq X^n_d + \frac{1}{m}$. If $t_i < s \leq b_i$ then $u_i(t_i, s, r_i(t_i, s)) - u_i(t_i, s, r_i(b_i, s)) = 0 - (t_i - s) \geq \frac{1}{m}$. Similarly, if $t_i \geq b_i + \frac{1}{m}$ then for $b_i < s \leq t_i$, $u_i(t_i, s, r_i(t_i, s)) - u_i(t_i, s, r_i(b_i, s)) = (t_i - s) - 0 \geq \frac{1}{m}$.

Let M_{dg} be a mechanism as in Corollary 2, where a Uniform Commitment mechanism is used:

Corollary 3 M_{dg} *$O(\sqrt{\frac{M^2}{N}\ln(N/2)})$-implements F in ex-post Nash equilibrium.*

This technique can extend to valuations and prices that are not restricted to the grid but can take any value in the unit interval. For this setting let M'_{dg} be the mechanism M_{dg} applied to the rounded announcements. That is we replace each announcement b_i with the the largest value in the grid S_M not exceeding it and apply M_{dg}. The loss of M'_{dg} is that of M_{dg} plus the effect of discretization, which adds to F at most $\frac{1}{M}$, and hence we get that M'_{dg} is an $(\sqrt{\frac{M^2}{N}\ln(N/2)} + \frac{1}{M})$-implementation. Setting $M = (N/\ln N)^{1/4}$ we get that M'_{dg} is an $O\left((\ln(N)/N)^{1/4}\right)$-implementation for F in ex-post Nash equilibrium.

The literature on optimal pricing in this setting has so far concentrated on the private values case. However, it has provided better approximations. In particular, Balcan et al. [5], using sampling techniques from Machine Learning, provide a mechanism that $O(\frac{1}{\sqrt{n}})$-implements the maximal revenue.

5 Facility Location

Consider a population of n agents located on the unit interval. An agent's location is private information and a social planner needs to locate K similar facilities in order to minimize the average distance agents travel to the nearest facility.[11] We assume each agent wants to minimize her distance to the facility that services her. In particular, this entails that values (and reactions) are private.

5.1 The Discrete Case

We first consider the discrete case where locations are restricted to a finite grid on the unit interval, $L = L(m) = \{0, \frac{1}{m}, \frac{2}{m}, \ldots, 1\}$. Using the notation of previous sections, let $T_i = L$, $S = L^K$, and let $R_i = L$. The utility of agent i is

$$u_i(t_i, s, r_i) = \begin{cases} -|t_i - r_i| & \text{if } r_i \in s, \\ -1 & \text{otherwise.} \end{cases}$$

Hence, $r_i(b_i, s)$ is the facility closest to the locations of the facility in s closest to b_i. Let $F(t, s) = \frac{1}{n}\sum_{i=1}^{n} u_i(t_i, s, r_i(t_i, s))$ be the social utility function, which is 1-sensitive (i.e., $d = 1$).

First, consider the uniform commitment mechanism $\hat{M}^U$, which is based on the uniform distribution over S for the commitment mechanism. Now consider the mechanism $\hat{M}_{LOC1}$, based on the uniform commitment mechanism, as in Corollary 2

Corollary 4 *$\hat{M}_{LOC1}$ $6\sqrt{\frac{m(m+1)^K}{n}}\sqrt{\ln\left(\frac{n}{2m}\right)}$- implements the optimal location.*

Proof: Note that $\gamma = \frac{1}{m}$, $|S| = (m+1)^K$ and the proof follows immediately from Corollary 2.

QED

Now consider an alternative commitment mechanism. Consider the distribution P, over $S = L^K$, which chooses uniformly among all the following alternatives - placing one facility in location $\frac{j}{m}$ and the remaining $K-1$ facilities in location $\frac{j+1}{m}$, where $j = 0, \ldots, m-1$. Note that for any i, any pair $b_i \neq t_i$ is separated by at least one alternative in this set. For this mechanism $\tilde{p} = \frac{1}{m}$. Now consider the mechanism $\hat{M}_{LOC2}$, based on the commitment mechanism, M^P:

Corollary 5 *$\hat{M}_{LOC2}$ $6\sqrt{\frac{m^2}{n}}\sqrt{\ln\left(\frac{n(m+1)^K}{2m^2}\right)}$-implements the optimal location.*

[11] For expositional reasons we restricting attention to the unit interval and to the average travel distance. Similar results can be obtained for other sets in $\mathbb{R}^2$ and other metrics, such as distance squared.

Proof: This is an immediate consequence of Theorem 2.

QED

The rate at which the two mechanism converges to zero, as the society grows, is similar. In addition, both mechanisms deteriorate as the grid size , m, grows. However the latter deteriorates at a substantially slower rate. This becomes important when we leverage the discrete case to analyze the continuous case in the next paragraph.

In fact, we can further improve our results for the facility location problem. To do so we revisit the bound on the loss from being truthful for the Exponential Mechanism. This bound is based on the sensitivity of F and in particular on the fact that $|F(t,s) - F((\hat{t}_i, t_{-i}), s)| \leq \frac{1}{n}$ (recall that $d = 1$). In fact, a stronger inequality holds here: $|F(t,s) - F((\hat{t}_i, t_{-i}), s)| \leq \frac{|t_i - \hat{t}_i|}{n}$. This, in turn, yields a better bound on the loss incurred by being truthful in the Exponential mechanism:

Lemma 6 (Analog of Corollary 1) *In the facility location problem, if $\epsilon \leq 1$ then for any i, any $b_i, t_i \in T_i$ and any $t_{-i} \in T_{-i}$,*

$$E_{M^{\frac{\epsilon}{2}}(t_i,t_{-i})}[u_i(t_i, s, r_i(t_i, s))] \geq E_{M^{\frac{\epsilon}{2}}(b_i,t_{-i})}[u_i(t_i, s, r_i(t_i, s))] - 2\epsilon|t_i - b_i|.$$

Proof: Note that

$$\begin{aligned}
|F(b_i, t_{-i}, s) - F(t_i, t_{-i}, s)| &= \frac{1}{n}|u_i(b_i, s, r_i(b_i, s)) - u_i(t_i, s, r_i(t_i, s))| \\
&\leq \frac{1}{n}\max_{x=b_i,t_i} |u_i(b_i, s, r_i(x, s)) - u_i(t_i, s, r_i(x, s))| \\
&= \frac{1}{n}\max_{x=b_i,t_i} \big||b_i - r_i(x,s)| - |t_i - r_i(x,s)|\big| \leq \frac{1}{n}|t_i - b_i|.
\end{aligned}$$

Plugging this into the definition of the Exponential Mechanism we get:

$$\begin{aligned}
E_{M^{\frac{\epsilon}{2}}(b_i,t_{-i})}[u_i(t_i, s, r_i(t_i, s)] &= \sum_{s\in S} u_i(t_i, s, r_i(t_i, s))M^{\frac{\epsilon}{2}}(b_i, t_{-i})(s) \\
&= \sum_{s\in S} u_i(t_i, s, r_i(t_i, s))\frac{e^{\frac{n\epsilon}{2}F(b_i,t_{-i},s)}}{\sum_{s'\in S} e^{\frac{n\epsilon}{2}F(b_i,t_{-i},s')}} \\
&\leq \sum_{s\in S} u_i(t_i, s, r_i(t_i, s))\frac{e^{\frac{n\epsilon}{2}\left(F(t_i,t_{-i},s)+\frac{|t_i-b_i|}{n}\right)}}{\sum_{s'\in S} e^{\frac{n\epsilon}{2}\left(F(t_i,t_{-i},s')-\frac{|t_i-b_i|}{n}\right)}} \\
&= e^{\epsilon|t_i-b_i|}\sum_{s\in S} u_i(t_i, s, r_i(t_i, s))M^{\frac{\epsilon}{2}}(t_i, t_{-i})(s) \\
&= e^{\epsilon|t_i-b_i|}E_{M^{\frac{\epsilon}{2}}(t_i,t_{-i})}[u_i(t_i, s, r_i(t_i, s)],
\end{aligned}$$

The proof is completed by noting that as $|t_i - b_i| \leq 1$ and $\epsilon < 1$, $e^{\epsilon|t_i - b_i|} \leq 1 + 2\epsilon|t_i - b_i|$.

QED

Assume that the grid size is some power of 2, say $m = 2^{\bar{m}}$. Consider an imposing mechanism induced by the following distribution P over the set S: First, choose a uniformly a random number $X \in \{1, 2, 3, \ldots, \bar{m}\}$. Then choose another random number, Y uniformly in $\{0, 1, 2, \ldots, X\}$. Now let s be the alternative where one facility is located at $\frac{Y}{X}$ and the other $K - 1$ facilities at $\frac{Y+1}{X}$.

Lemma 7 (Analog of Lemma 4) *For the grid size* $m = 2^{\bar{m}}$, $\forall b_i \neq t_i, t_{-i}$, $E_{M^{P}(t_i,t_{-i})}[u_i(t_i, s)] \geq E_{M^{P}(b_i,t_{-i})}[u_i(t_i, s)] + \frac{|t_i - b_i|}{2\bar{m}}$.

Proof: Consider the case $b_i < t_i$. Note that whenever $1/X < |t_i - b_i| \leq 2/X$ and $b_i \leq \frac{Y}{X} < b_i + \frac{1}{X}$ then for the resulting social alternative, s, the facility assigned to i whenever she announces b_i is located at $\frac{Y}{X}$, whereas the facility assigned to i whenever she announces t_i is at $\frac{Y+1}{X}$. Consequently, $u_i(t_i, s, r_i(t_i, s)) \geq u_i(t_i, s, r_i(b_i, s)) + \frac{1}{X} \geq u_i(t_i, s, r_i(b_i, s)) + \frac{(t_i - b_i)}{2}$. In words, for the specific choice of X and Y misreporting one's type leads to a loss of over $\frac{(t_i - b_i)}{2}$.

The probability of choosing such X and Y is $\sum_{x=\frac{1}{t_i - b_i}}^{\frac{2}{t_i - b_i}} \frac{1}{\bar{m}} \frac{1}{x} \geq \sum_{x=\frac{1}{t_i - b_i}}^{\frac{2}{t_i - b_i}} \frac{1}{\bar{m}} \frac{t_i - b_i}{2} = \frac{1}{\bar{m}}$.

Since the mechanism is imposing, for any choice of X and Y misreporting is not profitable. Therefore, the expected loss from misreporting exceeds $\frac{|t_i - b_i|}{2\bar{m}}$.

The case $t_i < b_i$ is resolved with similar arguments.

QED

As in the generic construction , let $\bar{M}_q^{\epsilon}(t) = (1 - q)M^{\frac{\epsilon}{2}}(t) + qM_P(t)$. If q satisfies $q \cdot \frac{|t_i - b_i|}{2 \log_2 m} \geq 2\epsilon|t_i - b_i|$ then $\bar{M}_q^{\epsilon}(t)$ is truthful. In particular this holds for $q = 4\epsilon \log_2 m$.

Set $\hat{\gamma} = \frac{1}{m \cdot 2 \log_2 m}$, $\epsilon = \sqrt{\frac{\hat{\gamma} d}{n}} \sqrt{\ln\left(\frac{n\hat{\gamma}}{2d}|S|\right)}$ and $q = \frac{2\epsilon}{\hat{\gamma}}$ and let $\hat{M}_{LOC3} = \bar{M}_q^{\epsilon}$ for those parameters.

Theorem 3 (analog of Theorem 2) $\hat{M}_{LOC3}$ $O\left(\sqrt{\frac{m \ln m}{n}} \sqrt{\ln(nm)}\right)$*-implements F for large enough* n.

The proof follows similar arguments as those in the proof of Theorem 2 and is therefore omitted.

The quality of the approximation, in terms of the population size, is the same as the previous two mechanisms. However, the last mechanism is superior in terms of the grid size. This becomes instrumental when we analyze the continuous case.

5.2 The Continuous Case

We now use the above result to construct a mechanism for the case where types are taken from the (continuous) unit interval. Consider the mechanism $\hat{M}_{LOC3}$ for the grid with m elements and set $\hat{M}_{LOC4}(t) = \hat{M}_{LOC3}(round(t))$, where m will be determined henceforth and $round(t)$ is the vector of elements on the grid that is closest to t. In words, the mechanism $\hat{M}_{LOC4}$ first rounds agents' announcements to the closest point on the grid, and then applies the mechanism $\hat{M}_{LOC3}$ for the discrete case.

Recall that $\hat{M}_{LOC3}$ is truthful for the discrete case. Consequently, if i is located at t_i and announces a location b_i such that both b_i and t_i are rounded to the same element of the grid then the utility of i is similar. This shows that $\hat{M}_{LOC4}$ is not truthful, however it has weakly dominant strategies. In addition, the mechanism's outcomes are the same for all weakly dominant strategies. Hence, in order to compute how well $\hat{M}_{LOC4}$ implements F in weakly dominant strategies it suffices to analyze the outcome of $\hat{M}_{LOC4}$ assuming agents are truthful.

The loss of $\hat{M}_{LOC4}$ is bounded by that of $\hat{M}_{LOC3}$ and an additional additive factor of $\frac{1}{m}$, which is a result of the rounding. Hence we get that M_{LOC4} is $O\left(\sqrt{\frac{m \ln m}{n}}\sqrt{\ln(nm)} + \frac{1}{m}\right)$-optimal.

We now set $m = n^{1/3}/(\ln n)^{2/3}$ and next Theorem follows immediately:

Theorem 4 *$\hat{M}_{LOC4}$ weakly-$O\left(\frac{\ln(n)}{n^{1/3}}\right)$-implements F.*

6 Discussion

In this section we discuss the importance of the two driving forces of the mechanism: differential privacy and imposition. In addition, we discuss some of the limitations of our results.

6.1 Disposing of Imposition

The driving force underlying the proposed mechanism is the combination of differential privacy on the one hand with imposition on the other hand. One natural question is to understand the importance of each of these two components separately.

McSherry and Talwar [19] focus on differential privacy and show that this property is sufficient to yield approximate implementation in ϵ-dominant strategies. Unfortunately, this result cannot be qualitatively improved. Clearly, one cannot hope for exact implementation. However, as demonstrated in the following example, differential privacy on its own cannot guarantee implementation with a stronger solution concept. In fact, in the following, we show the existence of dominant strategies that lead to inferior result, in conjunction with the existence of the ϵ-dominant strategies.

Example 1 *Consider a monopolist producing a digital goods who faces n buyers, each having a unit demand at a valuation that is either 0.5 or 1. The monopolist cannot distinguish among buyers and is restricted to choosing a price in the set $\{0.5, 1\}$. Assume the monopolist is interested in maximizing the average revenue per buyer (we assume that the marginal cost of production for digital goods is zero).*[12] *The optimal outcome for the auctioneer is $OPT(\bar{t}) = \frac{\max_{s \in \{0.5,1\}}(s \cdot |\{i : t_i \geq s\}|)}{n}$. Indeed, if the monopolist uses the appropriate exponential mechanism then it is ϵ-dominant for agents to announce their valuation truthfully, resulting in almost optimal revenues. However, one should note that the probability that the exponential mechanism will choose the lower of the two prices increases with the number buyers that announce 0.5. Hence, it is* **dominant** *for buyers to announce 0.5. This may lead to inferior results. In particular, whenever all agents value the good at 1 but announce 0.5 the mechanism will choose the price 0.5 with high probability, leading to an average revenue of 0.5, which is half the optimal average revenue.*

6.2 Disposing of Differential Privacy

It is intuitive to assume that our notion of imposition trivializes the results. One could argue that the possibility to detect untruthful agents, even if it only in hind sight, is sufficient to induce truthful announcements, by inflicting sufficient punishment. Our notion of imposition may be viewed as a severe punishment mechanism, where agents are free to choose any reaction but are severely punished if the reaction falls outside of the restricted set. Does such punishment scheme trivialize our results? The answer is no. In fact, the next example demonstrates that imposition is insufficient to induce truthfulness. The reason for this is that an agent may benefit by mis-reporting, even if she later is committed to a sub-optimal reaction (the one that complies with her report instead of the one that complies with her type).

Example 2 *Consider a digital goods pricing problem where agents valuations are either $1 + \mu$ or $2 + \mu$, and the possible prices are 1 and 2. Consider a mechanism which implements the 'optimal' price with probability one (instead of the corresponding smoothing induced by the exponential mechanism). Assume there are $2n + 1$ agents, of which n announced a valuation of $1 + \mu$ and another n announced $2 + \mu$. Consider the last agent, who is of type $2 + \mu$. If she announces her true type then with high probability the price will is set to 2 and her expected profit is slightly greater than μ. If, on the other hand, she misrepresents her self and announces $1 + \mu$ then with high probability the price is set to 1 and her expected profit is slightly less than 1.*

In this example, no agent has a large influence of the value of the social welfare function. However, there exits a possibility where a single agent will have a large influence on the actual choice of the social alternative. Therefore, truthfulness is not dominant, although the environment is private and imposition mechanism is in place. By adding the exponential mechanism to the framework the choice of social alternative is no longer a 'knife-edge' decision, thus making each agent less influential and restoring truthfulness as a dominant strategy.

[12] We consider the average revenue per buyer as the social welfare function, instead of the total revenue, in order to comply with the requirement that the value of the social welfare function is restricted to the unit interval.

6.3 Model Limitations

There are three overarching limitations to the technique we present: (1) The generic mechanism only works for social welfare problems which are not sensitive, (2) We consider settings where the reaction set of agents is rich enough, such that any pair of types can be separated by the optimal reaction on at least one social alternative; and (3) The size of the set of social alternatives cannot grow too fast as the set of agents grows.

Sensitive function: As discussed earlier in the paper many of the social welfare functions of interest are actually insensitive and comply with our requirements. Examples of insensitive functions are are revenue in monopoly setting, sum of agents'valuations, the Gini index and more. However there are important settings where our technique cannot be applied as the social welfare function is sensitive. One important example is that revenue maximization in a single unit auction. Clearly there are extreme settings where a change in a type of a single agent can drastically change the revenue outcome. E.g., consider the instance where all agents value the good at zero, resulting in a maximal revenue of zero. Compare this with a unilateral change in the valuation agent 1 from zero to one. This will change the maximal revenue from zero to one as well. However, even in this case the domain of type profiles (valuation profiles) that demonstrate sensitivity is quite small, and in some generic sense this will not happen.

Rich reaction set: The second limitation, is that we consider settings where the reaction set of agents is rich enough. In fact, what we need for the results to hold is that for any pair of types of an agent there exists some social alternative for which the optimal reaction for the first type is different than the optimal reaction for the second type. For example, in an auction setting, we require that for each pair of agents' valuations the auctioneer can propose a price such that one type will buy the good, while the other will refuse.

Small number of social alternatives: The approximation accuracy we achieve is proportional to $\sqrt{\frac{\ln(nS)}{n}}$. Therefore, if S grows exponentially as the number of agents grows the inaccuracy may not vanish. Two examples for such settings are matching problems, where each social alternative specifies the list of pairs, and multi unit auctions where the number of goods is proportional to the number of bidders. On the other hand, if the size of the set S grows polynomially with the number of agents then the limiting accuracy result still holds, although the rate of conversion could be slower.

6.4 Alternative mechanisms

The framework we presented combines a differentially private mechanism with an imposing one. Our general results refer to a 'universal' construction of an imposing mechanism (the uniform one), yet the specific examples we analyze demonstrate that imposing mechanisms that are tailor made to the specific setting can improve upon the results.

Similarly, it is not imperative to use the Exponential mechanism as for the first component, and other differentially-private mechanisms may be adequate. In fact, the literature on differential privacy provides various alternatives that may outperform the Exponential mechanism in sa specific context. Some example can be found in Dwork et al. [9], where the mechanism has a noisy component that is calibrated to global sensitivity, or in Nissim et al. [22] where a similar noisy component is calibrated to smooth sensitivity. The latter work also uses random sampling to achieve similar properties. To learn more the reader is referred to the recent survey of Dwork [8].

References

[1] Dilip Abreu and Hitoshi Matsushima. Virtual Implementation in Iteratively Undominated Strategies: Complete Information. *Econometrica*, Vol. 60, No. 5 (Sep., 1992), pp. 993-1008.

[2] N. I. Al-Najjar and Rann Smorodinsky. Pivotal Players and the Characterization of Influence. *Journal of Economic Theory*, 92, pp. 318-342, 2000.

[3] N. I. Al-Najjar and Rann Smorodinsky. The efficiency of competitive mechanisms under private information. *Journal of Economic Theory*, 137:383–403, 2007.

[4] Noga Alon, Michal Feldman, Ariel D. Procaccia, and Moshe Tennenholtz. Strategyproof approximation mechanisms for location on networks. *CoRR*, abs/0907.2049, 2009.

[5] Maria-Florina Balcan, Avrim Blum, Jason D. Hartline, and Yishay Mansour. Mechanism design via machine learning. In *FOCS*, pages 605–614. IEEE Computer Society, 2005.

[6] E. Clarke. Multipart pricing of public goods. *Public Choice*, 18:19–33, 1971.

[7] Cynthia Dwork. Differential privacy. In Michele Bugliesi, Bart Preneel, Vladimiro Sassone, and Ingo Wegener, editors, *ICALP (2)*, volume 4052 of *Lecture Notes in Computer Science*, pages 1–12. Springer, 2006.

[8] Cynthia Dwork. The differential privacy frontier (extended abstract). In Omer Reingold, editor, *TCC*, volume 5444 of *Lecture Notes in Computer Science*, pages 496–502. Springer, 2009.

[9] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In Shai Halevi and Tal Rabin, editors, *TCC*, volume 3876 of *Lecture Notes in Computer Science*, pages 265–284. Springer, 2006.

[10] Dimitris Fotakis and Christos Tzamo Winner-Imposing Strategyproof Mechanisms for Multiple Facility Location Games. (in WINE 2010)

[11] Drew Fudenberg, David. K. Levine and Wolfgang Pesendorfer. When Are Non-Anonymous Players Negligible. *Journal of Economic Theory*, 79: 46-71, 1998.

[12] A. Gibbard. Manipulation of voting schemes. *Econometrica*, 41:587–601, 1973.

[13] Andrew V. Goldberg, Jason D. Hartline, Anna R. Karlin, Michael Saks, and Andrew Wright Competitive auctions *Games and Economic Behavior (Mini Special Issue: Electronic Market Design)*, 55, Issue 2, Pages 242-269, 2006.

[14] T. Groves. Incentives in teams. *Econometrica*, 41:617–631, 1973.

[15] George J. Mailath and Andrew Postlewaite Asymmetric Information Bargaining Problems with Many Agents *Review of Economic Studies*, 57: 351-367, 1990. 7. R. Myerson, “Population Uncertainty and Poisson Gam

[16] A. Mas-Colell, M.D. Whinston, and J.R. Green. *Microeconomic Theory*. Oxford University Press, 1995.

[17] Richard McLean and Andrew Postlewaite. Informational Size and Incentive Compatibility. *Econometrica*, 70: 2421-2454, 2002.

[18] Richard McLean and Andrew Postlewaite. Informational Size and Incentive Compatibility with Aggregate Uncertainty. *Games and Economic Behavior*, 45(2): 410-433, 2003.

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

[20] H. Moulin. On strategy-proofness and single-peakedness. *Public Choice*, 35:437–455, 1980.

[21] Noam Nisan and Amir Ronen. Algorithmic mechanism design (extended abstract). In *STOC*, pages 129–140, 1999.

[22] Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. Smooth sensitivity and sampling in private data analysis. In David S. Johnson and Uriel Feige, editors, *STOC*, pages 75–84. ACM, 2007.

[23] Ariel D. Procaccia and Moshe Tennenholtz. Approximate mechanism design without money. In *ACM Conference on Electronic Commerce*, pages 177–186, 2009.

[24] Rafael Rob Pollution Claim Settlements Under Private Information. *Journal of Economic Theory*, 47: 307-333, 1989.

[25] Kevin Roberts. The characterization of implementable choice rules. In Jean-Jacques Laffont, editor, *Aggregation and Revelation of Preferences. Papers presented at the 1st European Summer Workshop of the Econometric Society*, pages 321–349. 1979.

[26] D. J. Roberts and Andrew Postlewaite The Incentives for Price Taking Behavior in Large Exchange Economies *Econometrica*, 44: 115-127, 1976.

[27] Aldo Rustichini, Mark A. Satterthwaite -and Steven R. Williams. Convergence to efficiency in a simple market with incomplete information. *Econometrica*, 62(1):1041–1063, 1994.

[28] Mark A. Satterthwaite. Stratey proofness and arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10:187–217, 1975.

[29] Mark A. Satterthwaite and Steven R. Williams. The rate of convergence to efficiency in the buyers bid double auction as the market becomes large. *Review of Economic Studies*, 56:477–498, 1989.

[30] J. Schummer and R. V. Vohra. Strategy-proof location on a network. *Journal of Economic Theory*, 104(2):405–428, 2004.

[31] J. Schummer and R. V. Vohra. Mechanism design without money. In N. Nisan, T. Roughgarden, É. Tardos, and V. Vazirani, editors, *Algorithmic Game Theory*, chapter 10. Cambridge University Press, 2007.

[32] J. Swinkels. Efficiency of large private value auctions. *Econometrica*, 69(1):37–68, 2001.

[33] W. Vickrey. Counterspeculations, auctions, and competitive sealed tenders. *Journal of Finance*, 16:15–27, 1961.

[34] Y. Wang P. Lu, X. Sun and Z. Allen Zhu. Asymptotically optimal strategy-proof mechanisms for two-facility games. In *ACM Conference on Electronic Commerce*, 2010.

A A Critical Example of the McSherry & Talwar Model