# Approximately Optimal Mechanism Design via Differential Privacy*

Kobbi Nissim† Rann Smorodinsky‡ Moshe Tennenholtz§

July 21, 2010
draft

**Abstract**

A social planner wants to optimize the choice of a social alternative vis-a-vis the players' private information. We introduce a generic mechanism, **[[ Moshe:** the sentence is unclear **]]** this is (almost) the only context such that almost optimal choice is guaranteed in **[[ Moshe:** say strictly? **]]** dominant strategies. In addition, this mechanism does not hinge on monetary transfers. We demonstrate the mechanism in two specific contexts – location problems and pricing of digital goods. The design of this mechanism involves a lottery between two mechanisms – With high probability we actuate a mechanism that makes players **[[ Moshe:** informationally small may be unclear **]]** informationally small while being approximately optimal. The informational smallness ensures players cannot profit much by misreporting their true type. With the complementary probability we actuate a 'punishment' mechanism **[[ Moshe:** I'm not sure that the term punishment is not misleading; people may think about something much less sophisticated **]]** that provides strong incentives for being truthful.

*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 [10].

†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 [9]) 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 location of a facility to serve the agents, etc. The quality of the outcome is measured by some social welfare function that the social planner typically intends to maximize. This function can be the sum of the agents' valuations for an outcome in a public good setting, 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 [7] and Satterthwaite [18]) 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 [23] and generalized by Clarke [3] and Groves [8], guarantees **[[ Moshe:** guarantee? **]]** that being truthful is a dominant strategy and the outcome is efficient. Unfortunately, Roberts [16] showed that a similar mechanism cannot be obtained for other social welfare functions. This cul-de-sac induced researchers to look into truthful and approximately efficient mechanisms and this topic became the main subject matter of the work on algorithmic mechanism design, initiated in a paper by Nisan and Ronen [12]. **[[ Moshe:** In fact, in AGT the reason is computational complexity; most work deal with sum of valuations **]]** It turns out that compromising efficiency 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 [15]).

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 [11] and Shummer and Vohra [20, 21]). However, no *general* techniques are known for designing mechanisms that are approximately optimal. **[[ Moshe:** Do people in econ grasp approximation? should we add some explanation? **]]** Consider the facility location problem, as an example, where the social planner needs to locate some facilities, based on agents' report of their location. This problem has won **[[ Moshe:** won? 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. [1], Procaccia and Tennenholtz [15] and Wang et al. [14]). 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, and we provide a generic mechanism that is truthful. Furthermore, if the function being implemented is insensitive [[ **Moshe:** Insensitive at this point may be confusing since the reader can believe we speak about equilibrium ]] to unilateral changes then the mechanism is almost optimal as the size of the population, $n$, grows. The accuracy rate obtained by our mechanism for bounded function [[ **Moshe:** functions? ]] is of the order of of $O(\sqrt{\frac{\ln n}{n}})$. The resulting mechanism does not resort to 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 $\epsilon$-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 [10].
- With vanishing probability we deploy a punishment mechanism, which is designed to provide strong incentives to players to be truthful and punishes them otherwise. [[ **Moshe:** Again, punishment might be interpreted as something trivial ]]. The disadvantage of this component is that it provides a poor approximation to the optimal outcome.

The combined mechanism turns out to be truthful in dominant strategies, and provides an excellent approximation of the optimal outcome.

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) Location problems, where [[ **Moshe:** the? a? ]] society needs to decide on the location of $K$ facilities. In this setting we focus on minimizing 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.

Our 'punishment mechanism' is based on the possibility of holding agents to their announcement. We do so by considering a model where agents must take an action (hereinafter 'reaction'), once the social alternative is determined, in order to exploit it. By restricting agents' actions to the optimal action for the announced type we ensure that agents have an incentive to be truthful. This modeling choice is not standard, as typically we do not assume such a reaction phase. We motivate our modeling choice with a few examples:

**Facility Location:** Consider a location problem, where the social planner is faced with the challenge of locating $K$ facilities. The vector of locations is the social alternative. The initial announcements refer to agents' locations and the reaction is the choice of facility, among the available $K$ (typically, the closest one). **[[ Moshe:** At this point it is quite vague who reacts; we may lose the reader or fail to convince him due to that **]]**

**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. In return agents must choose between two reactions - buy at the proposed price or forego the opportunity.

**Exchange Economy:** Consider the exchange economy – $n$ agents arrive with their own bundle of goods (production bundle). The social planner determines a price vector, based on the total supply and the demand (which is private information). Then agents buy and sell according to the price vector. In this example the initial action could be an announcement of a demand function, whereas the 'reactions' are the amounts an agent actually trades once a price vector is determined.

**Public Good:** Consider a welfare problem **[[ Moshe:** unclear statement **]]** the central planner offers a set 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 location problem. **[[ Moshe:** Seems too vague. **]]**

**Network Design:** As a final example consider the problem of designing a communication or transportation network, based on agents' (privately known) needs. The 'reaction' in this case is the specific choice of vertices **[[ Moshe:** choice of vertices might be unclear **]]** chosen by the agent, once the network is built.

All these examples demonstrate the prevalence of 'reactions' in a typical design problem. In an abstract sense, the choice of reactions, can be thought of as part of the joint decision (the social planner's choice), sometimes referred to as the 'allocation'. We prefer to separate between the social alternative which is an element that is common to all the agents (e.g., the network, the price, the facility location) and the private components. The mechanism we construct typically (with high probability) does not assign an allocation, and provides full flexibility to the players in choosing their reaction. In particular, we provide a detailed analysis of the first two examples.

## 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 efficiency when the population of players is large. Some examples of work that hinge on a similar principle for large, yet finite populations, are Swinkels [22] who studies auctions, Satterthwaite and Williams [19] and Rustichini, Satterthwaite and Williams [17] who study double auctions, and Al-Najjar and Smorodinsky [**?**] 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 [**?**] who study an exchange economy).

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

**[[ Moshe:** At some point we need to make a claim of what is novel in ours comparing to the others above **]]**

**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 [6] and Dwork [4], 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 [5] for a recent survey). In this strand of the literature computations (equivalently mechanisms) that preserve differential privacy are referred to as $\epsilon$-*differentially private* computations.

McSherry and Talwar [10] establish an inspiring connection between differential privacy and mechanism design. They observe that participants (players) that contribute private information to $\epsilon$-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 $\epsilon$-differentially private truth-telling is approximately dominant, regardless of the agent utility functions. McSherry and Talwar introduce the exponential mechanism as a general technique for constructing almost optimal mechanisms that are almost incentive-compatible. **[[ Moshe:** Notice the use of incentive compatible vs. truthful, and also whether we talk about strict or weak dominance, where the latter is more common **]]** They 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 $\epsilon$-dominant for the exponential mechanism. On the one hand, lower values of $\epsilon$ imply higher compatibility with incentives. On the other hand, lower values deteriorate the approximation results. What is the optimal choice for $\epsilon$? 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.

In the economic community some attempts to formalize an abstract notion of influence have also been made. Fudenberg **[[ Moshe:** missing ",". Also it is unclear how this defers from ours; seems too much like a "reshimat kvisa" when the context is unclear **]]** Levine and Pesendorfer [**?**] and Al-Najar and Smorodinsky [**?**] provide one such attempt to formalize influence and prove bounds on average influence and on the number of influential players. McLean and Postlewaite [**?**, **?**] 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 [14] 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.[1]

**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 in this

[1]The notation $\tilde{O}(\cdot)$ is a convention in the computer sciences literature. **[[ Moshe:** Check this and I think there should be nothing here special to CS.... **]]** A function $f : N \to \mathbb{R}$ gives a $\tilde{O}(n^{-1})$ approximation to the function $g : N \to \mathbb{R}$ if $\frac{n|f(n)-g(n)|}{\ln(n)} \to_{n\to\infty} 0$.

settings have been studies by Goldberg et al [**?**] and Balcan et al [2]. Balcan et al [2] 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. However, the mechanism we propose is an instance of a general theory on mechanism design and is not ad-hoc. **[[ Moshe:** better statement is needed. also, if there is some aspect in which our mechanism is better, e.g. requires less info, then this can help **]]**

# 2 Model

Let $N$ denote the **[[ Moshe:** the? 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 $A_i$ be the set of reactions available to $i$. Typically, once a social alternative, $s \in S$, is determined agents choose a reaction $a_i \in A_i$. The utility of an agent $i$ is therefore a function of her type, the chosen social alternative and the chosen reaction. Formally, $u_i : T_i \times S \times A_i \to [0, 1]$.[2] A tuple $(T, S, A, u)$, where $A = \times_{i=1}^{n} A_i$ and $u = (u_1, \ldots, u_n)$, is called an *environment*.

Consider the set $\textit{argmax}_{a_i \in A_i} u_i(t_i, s, a_i)$, consisting of the optimal reactions available to $i$, at type $t_i$ and alternative $s$. We will sometimes abuse notation and denote by $a_i(t_i, s)$ an arbitrary optimal reaction (i.e., $a_i(t_i, s)$ is an arbitrary function which image is in $\textit{argmax}_{a_i \in A_i} u_i(t_i, s, a_i)$).

We say that an environment is *non-trivial* if for any $i$ and $\forall t_i \neq \hat{t}_i \in T_i$ there exists some $s \in S$, denoted $s(t_i, \hat{t}_i)$, such that $\textit{argmax}_{a_i \in A_i} u_i(t_i, s, a_i) \cap \textit{argmax}_{a_i \in A_i} u_i(\hat{t}_i, s, a_i) = \emptyset$. We say that $s(t_i, \hat{t}_i)$ *separates* between $t_i$ and $\hat{t}_i$. Let $\tilde{S} = \{s(t_i, \hat{t}_i) | i \in N, t_i \neq \hat{t}_i \in T_i\} \subset S$ be the set of social alternatives that separate any two types. **[[ Moshe:** do we need any element here to separate between any two types, or something simpler; seems vague at this point **]]**

A social planner, not knowing the vector of types, wants to maximize an arbitrary *social welfare function*, $F : T \times S \to [0, 1]$.[3] 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.*

Two examples of 1-sensitive functions are the average utility, $F = \frac{\sum_i u_i}{n}$, and the Gini coefficient. Note that a $d$-sensitive function prevents situations where any single agent has an overwhelming impact on the social choice. In fact, if a social welfare function is not $d$-sensitive, for

[2] 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.

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

any $d$, then in a large population this function is susceptible to minor faults in the system (e.g., noisy communication channels).[4] [[ **Moshe:** Someone might see the restriction to functions which are not sensitive as trivial when we deal with epsilon deviations; I guess we need to explain why something like DP is not trivial ]]

Denote by $\mathcal{A}_i = 2^{A_i} \setminus \{\emptyset\}$ the set of all subsets of $A_i$, except for the empty set, and let $\mathcal{A} = \times_i \hat{A}_i$. [[ **Moshe:** Something unclear or wrong above ]]

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

In words, for any $t$, $M(t)$ is a probability distribution over the set of social alternative and the agents' set of possible reactions. So a mechanism, in fact, randomly chooses a social alternative as well as a restricted subset of reactions for each agent $i$. Let $M_S(t)$ denote the marginal distribution of $M(t)$ on $S$ and let $M_i(t)$ denote the marginal distribution on $\mathcal{A}_i$. If the grand set of reactions, $A_i$, is always chosen by the mechanism then we say it is *non-imposing*. More formally, if $M_i(t)(A_i) = 1$ (the probability assigned to the grand set of reactions is one), for all $i$ and $t \in T$ then we say that $M$ is *non-imposing*. $M$ is *$\epsilon$-imposing* if $M_i(t)(A_i) \geq 1 - \epsilon$ for all $i$ and $t \in T$. [[ **Moshe:** The whole thing about imposition is quite cumbersome; I'm not sure what we should do; I was almost losing it again while reading.... ]]

A strategy of an agent is a choice of announcement, given her type, and a choice of reaction once the social alternative and her subset of reactions have been determined by the mechanism. However, for equilibrium analysis, we can restrict attention to dominant strategies and therefore we will assume that once a social alternative is chosen the agent will choose some reaction that maximizes her utility from the set of allowable reactions, that is for any $t_i$, $s$ and $\hat{A}_i \subset A_i$ agent $i$ chooses some $a_i(t_i, s) \in \textit{argmax}_{a_i \in A_i} u_i(t_i, s, a_i)$. As one agent's utility is independent of the choice of reactions by other agents the equilibrium analysis is independent of which such optimal reaction is taken. Therefore, formally, a strategy for $i$ can be reduced to function $W : T_i \to T_i$.

Given 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 the available reaction sets, $S \times \mathcal{A}$. The expected utility of $i$, at a vector of types $t$, is $E_{M(W(t))}[\max_{a_i \in \hat{A}_i} u_i(t_i, s, a_i)]$.

A strategy $W$ is *dominant* for the mechanism $M$ if for any $i$, $t_i$, $b_{-i}$ and $b_i \neq W_i(t_i)$ the following holds: $E_{M(W_i(t_i), b_{-i})}[\max_{a_i \in \hat{A}_i} u_i(t_i, s, a_i)] > E_{M(b_i, b_{-i})}[\max_{a_i \in \hat{A}_i} u_i(t_i, s, a_i)]$. In words, the expected utility for $i$ from complying with the strategy $W_i$ is greater than the utility from announcing some other type $b_i$, no matter what other agents announce. If the strategy $W_i(t_i) = t_i$ is dominant for all $i$ then $M$ is *truthful* (or *incentive compatible*) and $M(W(t)) = M(t)$.

[4] Most examples of social welfare function studies in the literature are $d$-sensitive. [[ **Moshe:** for which d? do we assume constant d? ]] One example of a function that is not $d$-sensitive is the parity function where $F = 1$ if the number of agents whose utility exceeds some threshold is odd, and $F = 0$ otherwise.

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

We say that the mechanism $M$ *implements* the social function $F$ if there exists a dominant strategy tuple, $W$, such that for all $t$, $E_{M(W(t))}(F(t,s)) = \max_{s \in S} F(t,s)$. **[[ Moshe:** Here is a point that the reader may lose trick, since we don't refer to reactions; it is a bit subtle set of definition **]]** It is generally impossible to find a mechanism that implements some abstract function, $F$. However, approximate implementation, in a sense made accurate below, is possible.

**Definition 3 ($\beta$-implementation)** *A mechanism $M$* $\beta$-implements *$F$, for $\beta > 0$, if there exists some dominant strategy tuple, $W$, such that for any $t \in T$, $E_{M((W(t))}(F(t,s)) \geq max_{s \in S} F(t,s) - \beta$.*

Our main result, stated informally, is

**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 $\beta$-implements $F$ for populations with more than $n_0$ agents.

**[[ Kobbi:** in the previous theorem, truthtelling is a dominant strategy, no? **]]**

**[[ Rann:** yes, but its not where the meat of the result lies and so I did not mention it **]]**

In addition to a generic mechanism we study two specific models and cast our generic results onto those settings. In the two models, facility location and pricing, we also provide implementation when the type sets and the set of social alternatives are large. **[[ Moshe:** large? infinite? **]]** In these extensions we use the following solution concept:

**Definition 4 (Weak-Dominance)** *A strategy $W_i$ is* weakly dominant *for the mechanism $M$ if for any, $t_i$, $b_{-i}$ and $b_i \neq W_i(t_i)$ $E_{M(W_i(t_i),b_{-i})}[u_i(t_i,s)] \geq E_{M(b_i,b_{-i})}[u_i(t_i,s)]$. A strategy tuple $W$ is* weakly dominant *if $W_i$ is weakly dominant for each $i$.*[5]

**[[ Kobbi:** don't we get dominance in the continuous case? **]]**

**[[ Rann:** due to the rounding we do not... or do we? let me think about it **]]**

Note that we do not require a weakly dominant strategy to be strictly better than other strategies on some realization.

**Definition 5** *The mechanism $M$* weakly $\beta$-implements *$F$, for $\beta > 0$, if the set of weakly dominant strategy tuples is not empty, and for any $W$ in that set and every $t \in T$, $E_{M((W(t))}(F(t,s)) \geq max_{\bar{s}} F(t,\hat{s}) - \beta$.*

[5]Note that a weak inequality replaces the strict inequality from the definition of dominant strategies. Also note that weak dominance does not require a strong inequality for at least one instance.

# 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 Punishment mechanism which is imposing but ignores agents' announcements when (randomly) selecting a social alternative.

## 3.1 The Exponential Mechanism

Consider the following non-imposing mechanism, which we refer to as the Exponential mechanism, originally introduced by McSherry and Talwar [10]:

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

[[ **Kobbi:** The above is for the discrete case. do we want also to have the continuous version? ]]

[[ **Rann:** do we need the continuous case in the sequel? I think not. If so, then no need to complicate things ]]

The Exponential mechanism has two notable properties, as we show below. The first property is that of $\epsilon$-differential privacy, which is inspired by the literature on privacy. We follow Dwork et al. [6] and define:

**Definition 6 ($\epsilon$-differential privacy)** *A mechanism, $M$, provides* $\epsilon$-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, $M(t)(s) \leq e^{\epsilon} \cdot M(\hat{t})(s)$.*[6]

In words, a mechanism preserves $\epsilon$-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 $e^{\epsilon}$, which approaches 1 as $\epsilon$ approaches zero.[7]

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

[7] The motivation underlying this definition of $\epsilon$-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 1 (McSherry and Talwar [10])** *If $F$ is $d$-sensitive then $M^{\frac{\epsilon}{2d}}(t)$ preserves $\epsilon$-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 $s \in S$:

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

**QED**

In addition, McSherry and Talwar [10] observe that, in the context of mechanism design, if a mechanism provides $\epsilon$-differential privacy then it is almost dominant for players to be truthful in the following sense:

**Lemma 2** *If $M$ in non-imposing and provides $\epsilon$-differential privacy, where $\epsilon < 1$, then for any $i$, any $b_i, t_i \in T_i$ and any $t_{-i} \in T_{-i}$,*

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

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

Combining Lemmas 1 and 2 we derive the first property of the Exponential mechanism:

**Corollary 1 (McSherry and Talwar [10])** *If $F$ is $d$-sensitive then for the mechanism $M^{\frac{\epsilon}{2d}}(t)$ agents can profit no more than $2\epsilon$ by misreporting their true type.*

In fact, a closer look at the derivations above shows that agents are $2\epsilon$-indifferent between any two strategies **[[ Kobbi:** shall we say that this was ignored by MT? **]] [[ Rann:** I think we should not position the paper as a 'reply' to MT but as an independent paper and therefore I think we should not add this comment **]]**. Another important property of the Exponential mechanism, due to McSherry and Talwar [10], is that whenever agents are truthful the outcome is nearly optimal:

**Lemma 3 (McSherry and Talwar [10])** *Let $F : T^n \times S \to [0, 1]$ be an arbitrary $d$-sensitive social welfare function and $n > \frac{e2d}{\epsilon|S|}$. Then for any $t$, $E_{M^{\frac{\epsilon}{2d}}(t)}[F(t, s)] \geq \max_s F(t, s) - \frac{4d}{n\epsilon} \ln\left(\frac{n\epsilon|S|}{2d}\right)$.*

Note that the $\lim_{n\to\infty} \frac{4d}{n\epsilon} \ln\left(\frac{n\epsilon|S|}{2d}\right) = 0$. Therefore, the exponential mechanism is almost optimal for a large and truthful population. The proof is standard, and is provided to completeness:

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

**[[ Moshe:** the max below should be over the $\hat{s}$ **]]** Fix a vector of types, $t$ and denote by $\hat{S} = \{s \in S : F(t,\hat{s}) < \max_s F(t,s) - \delta\}$. For any $s \in \hat{S}$ 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 $\delta$ 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: As shown the Exponential Mechanism has two properties - '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. [6], the addition of noise calibrated to smooth sensitivity and the sample and aggregate framework by Nissim et al. [13]. The reader is further referred to the recent survey of Dwork [5].

## 3.2 The Punishment Mechanism

We now consider an imposing mechanism that chooses $s \in S$ randomly, while ignoring players announcements. Once $s$ is chosen the mechanism restricts the allowable reactions for $i$ to

those that are optimal assuming she was truthful. Formally, if $s$ is chosen according to a the probability distribution $P$, let $M^P$ denote the following mechanism: $M^P_S(t)(s) = P(s)$ and $M^U_i(t)(a_i(t_i, s))|s) = 1$. Players do not influence the choice of $s$ in $M^U$ and so they are (weakly) better off being truthful. **[[ Moshe:** I hope people will feel fine with the fact we add such super-scripts to $M$ without too much discussion **]]**

**[[ Kobbi:** Do we have both $M^P$ and $M^U$, or should everything be $M^P$? **]]**

**[[ Rann:** $M^U$ is an instance of $M^P$ for the case $P = U$ **]]**

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

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

In words, $\gamma$ is a lower bound for the loss incurred by misreporting in case of an adversarial choice of $s \in S$. Recall the set of separating alternatives, $\tilde{S}$. We say the a distribution $P$ is *separating* if it assigns a positive probability to any element 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* **[[ Moshe:** *is non-trivial clear? is separating clear here or we need to re-visit?* **]]** *and $P$ is a separating distribution over $S$ then $\forall b_i \neq t_i, t_{-i}$, $E_{M^P(t_i, t_{-i})}[u_i(t_i, s)] \geq E_{M^P(b_i, t_{-i})}[u_i(t_i, s)] + \tilde{p}\gamma$ , and in particular truthfulness is a dominant strategy for all agents.*

**Proof**: For any pair $b_i \neq t_i$ there exists some $s = s(t_i, b_i)$ for which **[[ Moshe:** missing ")" **]]** $u_i(t_i, s, a_i(t_i, s)) \geq u_i(t_i, s, a_i(b_i, s) + \gamma$. $P$ 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.

## 3.3 A Generic and Nearly Optimal Mechanism

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

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

**Theorem 1** *If $q\gamma \geq \frac{2\epsilon}{\tilde{p}}$ then $\bar{M}^\epsilon_q$ is truthful.*

**Proof**: Follows immediately from lemmas 2 and 4. **[[ Moshe:** perhaps we should put in the numbers; we have done it in other straightforward cases **]]** **QED**

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

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

and consider populations of size $n > n_0$, where $n_0 = \max\{\frac{2\tilde{p}d}{\gamma}\ln\left(\frac{\gamma|S|}{2d}\right), \frac{4e^2 d}{\tilde{p}\gamma|S|^2}\}$ and in addition $\frac{n_0}{\ln(n_0)} > \frac{2\tilde{p}d}{\gamma}$.

**Lemma 5** *If $n > n_0$ then*

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

**Proof**: Part (1) follows from part (2), as $\tilde{p} < 1$.

Part (2): $\frac{n}{\ln(n)} > \frac{n_0}{\ln(n_0)} > \frac{2\tilde{p}d}{\gamma}$ which implies $n > \frac{2\tilde{p}d}{\gamma}\ln(n)$, and also $n > n_0 > \frac{2\tilde{p}d}{\gamma}\ln\left(\frac{\gamma|S|}{2d}\right)$. [[ **Moshe:** Isn't the last one inequality? ]] Therefore $n > \frac{\tilde{p}d}{\gamma}\ln\left(\frac{\gamma|S|}{2d}\right) + \frac{\tilde{p}d}{\gamma}\ln(n) = \frac{\tilde{p}d}{\gamma}\ln\left(\frac{\gamma|S|n}{2d}\right) \implies \gamma^2 > \frac{\gamma\tilde{p}d}{n}\ln\left(\frac{\gamma|S|n}{2d}\right)$. Taking the square root yields the desired result.

Part (3): [[ **Moshe:** Missed something in the strict inequality following $n_0$ ]] $n > n_0 > \frac{4e^2 d}{\gamma|S|} > \frac{4e^2 d}{\tilde{p}\gamma|S|^2} \implies \sqrt{n} > \frac{2ed}{\sqrt{\tilde{p}\gamma d}|S|}$. In additions $n > \frac{4e^2 d}{\gamma|S|} > \frac{2de}{\gamma|S|}$ which implies $1 < \ln\left(\frac{\gamma|S|n}{2d}\right)$. Combining these two inequalities we get: $\sqrt{n} > \frac{2ed}{\sqrt{\tilde{p}\gamma d}\sqrt{\ln\left(\frac{\gamma|S|n}{2d}\right)}|S|}$. Multiplying both sides by $\sqrt{n}$ implies $n > \frac{2ed\sqrt{n}}{\sqrt{\tilde{p}\gamma d}\sqrt{\ln\left(\frac{\gamma|S|n}{2d}\right)}|S|} = \frac{2ed}{\epsilon|S|}$.

**QED**

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

**Theorem 2 (Main Theorem)** *The mechanism $\hat{M}(t)$ $6\sqrt{\frac{d}{\tilde{p}\gamma n}}\sqrt{\ln\left(\frac{n\gamma|S|}{2d}\right)}$-implements $F$, for $n > n_0$.*

**Proof**: Given the choice of parameters $\epsilon$ and $q$ then, Theorem 1 guarantees that $\hat{M}(t)$ is truthful. Therefore, it is sufficient to show that for any type vector $t$,

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

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 **[[ Moshe:** make this more explicit? **]]** 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

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

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

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

**[[ Moshe:** perhaps the fact of why $|S|$ disappears at he last inequality should be explained; I guess it is just since of its size **]]** and the result follows.

**QED**

One particular case of interests is the punishment mechanism $M^U$, where $U$ is the uniform distribution over the set $S$. Let $\hat{M}^U$ denote the mechanism from Theorem 2 for the instance $P = U$. In this case $\tilde{p} = \frac{1}{|S|}$ and therefore:

**Corollary 2** *The mechanism* $\hat{M}^U(t) \quad 6\sqrt{\frac{d|S|}{\gamma n}}\sqrt{\ln\left(\frac{n\gamma|S|}{2d}\right)}$*-implements* $F$, $\forall n > n_0$.

**[[ Moshe:** I would suggest to expand here; the analysis is simple, but that's a qualitative different case, and wish to emphasize **]]**

Both the result in the main theorem and the one in the corollary exhibit convergence 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. We take advantage of this observation in the analysis of two applications.

# 4 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.[8]

## 4.1 The Discrete Case

We first consider the discrete case where locations are restricted to 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 $A_i = L$. The utility of agent $i$ is

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

In addition, let $F(t, s) = \frac{1}{n}\sum_{i=1}^{n} u_i$ be the social utility function, which is 1-sensitive (i.e., $d = 1$). The set of reactions $A_i$, is the set of facility locations, and $a_i(b_i, s)$ is the facility closest to the locations of the facility in $s$ closest **[[ Moshe:** something in the statement is unclear; closest closest... **]]** to $b_i$. Clearly, $F$ is 1-sensitive and $|S| = m^K$. **[[ Moshe:** Aren't there m+1 locations? **]]**

First, let consider the uniform punisg=hment **[[ Moshe:** typo **]]** mechanism $\hat{M}^U$, which is based on the uniform distribution over $S$ for the punishment mechanism. Now consider the mechanism $\hat{M}_{LOC1}$, based on the uniform punishment mechanism, as in Corollary 2

**Corollary 3** $\hat{M}_{LOC1}$ $6\sqrt{\frac{m^{K+1}}{n}}\sqrt{\ln\left(\frac{nm^{K-1}}{2}\right)}$- *implements the optimal location.*

**Proof**: Note that $\gamma = \frac{1}{m}$ and the proof follows immediately from Corollary 2.

**[[ Moshe:** Perhaps we should put the numbers; somewhat hard to check **]]** **QED**

Now let us consider an alternative punishment 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 pair of $b_i \neq t_i$ 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 punishment mechanism, $M^P$:

[8]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.

**Corollary 4** $\hat{M}_{LOC2}$ $6\sqrt{\frac{m^2}{n}}\sqrt{\ln\left(\frac{nm^{K-1}}{2}\right)}$*-implements the optimal location.*

**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, in the next paragraph.

In fact, we can further improve our results for the facility location. 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 $d = 1$). However, a unilateral change in one of the agents' locations, say from $t_i$ to $\hat{t}_i$, changes the average travel distance by no more than $\frac{|t_i - \hat{t}_i|}{n}$. Formally, $|F(t,s) - F((\hat{t}_i, t_{-i}), s)| \leq \frac{|t_i - \hat{t}_i|}{n}$.

This, in turn, yields a better bound on the loss incurred by being truthful in Exponential mechanism. If $i$'s location is $t_i$ then she can profit no more than $2\epsilon|b_i - t_i|$ by reporting $b_i$ to the Exponential mechanism:

**Lemma 6 (1)** *For the location problem with $K = 2$, for any $i$, any $b_i \neq t_i \in T_i$ and any $t_{-i} \in T_{-i}$,*

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

To leverage this we consider a punishment mechanism that 'guesses' by how much an agent misreports her type and provides an optimal punishment level. For the sake of simplicity we pursue this idea for the case $K = 2$ and restrict attention to the case where $m$ is even.

Formally, consider the following distribution $P$ for the Punishment mechanism: (1) Choose a random number $X \in \{1, 2, 4, \ldots, \log_2 m\}$ uniformly and set $\Delta = \frac{X}{m}$. (2) Consider the set of pairs of facility locations of the form $\{s, s + \Delta/2\}$ and choose one such pair randomly and uniformly.

**Lemma 7 (Analog of Lemma 4)** $\forall b_i \neq t_i, t_{-i}$, $E_{M^P(t_i, t_{-i})}[u_i(t_i, s)] \geq E_{M^P(b_i, t_{-i})}[u_i(t_i, s)] + ??$

**[[ Moshe:** I'm confused by the ??? **]]** **Proof:** Consider the case $b_i < t_i$. If $\Delta$ is satisfies $(t_i - b_i)/2 \leq \Delta < t_i - b_i$ and $s$ satisfies $t_i \leq s < t_i + \Delta/2$ then $u_i(t_i, s, a_i(t_i, s)) \geq u_i(t_i, s, a_i(b_i, s)) + \Delta/2 \geq u_i(t_i, s, a_i(b_i, s)) + (t_i - b_i)/2$. In words, misreporting one's type leads to a loss of over $(t_i - b_i)/2$.

The probability for choosing such $\Delta$ and $s$ is $\sum_{\delta=\frac{t_i-b_i}{2}}^{t_i-b_i} \frac{1}{\log_2 m}\frac{2}{\delta} \geq \sum_{\delta=\frac{t_i-b_i}{2}}^{t_i-b_i} \frac{1}{\log_2 m}\frac{2}{t_i-b_i} = \frac{1}{\log_2 m}$.

For any other $\Delta$ and $s$ misreporting is not profitable and so the expected loss from misreporting exceeds $\frac{|t_i-b_i|}{2\log_2 m}$.

Using the same arguments yields the same inequality for the case $t_i < b_i$.

**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 provided in the appendix. ??

**[[ Rann:** We should write explicit proofs for the claims above, which should go into an appendix. Kobbi - can work it out? **]]**

**[[ Moshe:** The claim above should have proofs; also we should explain the aim of these; what we hope to do better/why/when **]]** 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.

## 4.2 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 with $m = n^{1/3}/(\ln n)^{2/3}$ elements and set $\hat{M}_{LOC4}(t) = \hat{M}_{LOC3}(round(t))$, where $round(t)$ is the vector of elements on the grid that closest to $t$. In words, $\hat{M}_{LOC4}$ first rounds agents' announcements to the closest point on the grid, and then applies the mechanism for the discrete case.

Recall that $\hat{M}_{LOC3}$ is truthful for for the discrete case. Consequently, if $i$ is located at $t_i$ and announces a location $b_i$ 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. Substituting $m$ for $n^{1/3}/(\ln n)^{2/3}$ provides a proof for the following:

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

# 5 Pricing Digital Goods

A monopolist that produces digital goods, for which the marginal cost of production is zero, faces a set of $n$ indistinguishable buyers. Each buyer has a unit demand with a privately known reservation price (her valuation). The monopolist wants to set a price in order to maximize her revenue.

**The Discrete Case.** We cast this problem into our framework and, as with the previous example, begin with the discrete case. Let $S = \{0, \frac{1}{m}, \frac{2}{m}, \ldots, 1\}$ be the set of possible valuations and prices. Let set of reactions for $i$ be $A_i = \{$'Buy', 'Not buy'$\}$, and so the utility function for agent $i$ is

$$u_i(t_i, s, a_i) = \begin{cases} t_i - s & \text{if } a_i = \text{'Buy'}, \\ 0 & \text{if } a_i = \text{'Not buy'}. \end{cases}$$

Let $F(t, s) = \frac{s}{n} \cdot |\{i : t_i \geq s\}|$ be the social welfare function (the monopolists's profit) and note it is 1-sensitive.

In this application the gap is equal $\frac{1}{m}$. To see this consider the case $b_i \geq t_i + \frac{1}{m}$. If $t_i < s \leq b_i$ then $u_i(t_i, s, a_i(t_i, s)) - u_i(t_i, s, a_i(b_i, s)) = 0 - (t_i - s) \geq \frac{1}{m}$. Similarly, if $t_i \geq b_i + \frac{1}{m}$ then for $b_i < s \leq t_i$, $u_i(t_i, s, a_i(t_i, s)) - u_i(t_i, s, a_i(b_i, s)) = (t_i - s) - 0 \geq \frac{1}{m}$.

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

**Corollary 5** *$M_{dg}$ $O(\sqrt{\frac{m^2}{n}\ln(n/2)})$-implements $F$.*

This result can be improved to an $(\ln(n))^{2/3}/n^{1/3}$-implementation using a similar technique to that presented for 2-median above.

**[[ Rann:** Kobbi - can you work this out in details? **]]**

**The Continuous Case.** As with the previous example, we can modify the construction of $M_{dg}$ to the case where prices and agent valuations are taken from the interval $[0, 1]$. The new mechanism $M'_{dg}$ first rounds each of its inputs $b_i$ to the largest value in $L$ not exceeding it, and then applies $M_{dg}$. 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 $?? + \frac{1}{m}$)-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 $F$.

As mentioned in the intrduction, ad-hoc mechanism for pricing can achieve better results than our generic technique. In particular, Balcan et al. [2], using sampling techniques from Machine Learning, provide a mechanism that $O(\frac{1}{\sqrt{n}})$-implements the maximal revenue.

# References


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

[2] 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.

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

[4] 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.

[5] 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.

[6] 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.

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

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

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

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

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

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

[13] 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.

[14] 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.

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

[16] 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.

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

[18] M.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.

[19] 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.

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

[21] 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.

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

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

# A A Counter Example to the MT-Paradigm

The following is an example of differentially private mechanism, a-la McSherry and Talwar, in the Digital Goods auction setting. The interesting aspect of this example is that the mechanism has a unique type-independent dominant strategy equilibrium, where agents are not truthful, which leads to very low revenue.

**Example 1 (Digital Goods (simplified))** *Consider a model of $n$ agents in an auction for digital goods, with unit demand, and assume each agent values are taken from the type set $T = \{0.5, 1\}$. We will also restrict the alternative set $S$ to be $\{0.5, 1\}$. The optimal revenue for the auctioneer is hence $OPT(\bar{t}) = \max_{s\in\{0.5,1\}}(s \cdot |\{i : t_i \geq s\}|)$. Construct the exponential mechanism with $q(\bar{b}, s) = s \cdot |\{i : b_i \geq s\}|$. Note that $\Delta q \leq 1$, and hence the result is $2\epsilon$-differentially private. McSherry and Talwar showed that if agents are truthful (i.e., $\bar{b} = \bar{t}$), then the expected revenue of this mechanism is $OPT(\bar{t}) - O((\log n)/\epsilon)$.*

*Let us have a closer look at the mechanism. On agents bids $\bar{b}$, the mechanism outputs $0.5$ with probability*

$$\frac{\exp(\epsilon \cdot 0.5 \cdot |\{i : b_i \geq 0.5\}|)}{\exp(\epsilon \cdot 0.5 \cdot |\{i : b_i \geq 0.5\}|) + \exp(\epsilon \cdot |\{i : b_i = 1\}|)} = \frac{exp(\epsilon n/2)}{exp(\epsilon n/2) + \exp(\epsilon \cdot |\{i : b_i = 1\}|)},$$

*and otherwise it outputs 1. It is quite straightforward to see that each agent will strictly benefit from bidding a price of $0.5$ over bidding $1$, no matter what the others bid, as that would increase the probability that the mechanism will choose the price $0.5$ over the price $1$.*

*Thus, for all agents the* unique dominant strategy *is to bid $0.5$, leading to a price choice of $0.5$ with high probability $\frac{exp(\epsilon n/2)}{exp(\epsilon n/2)+1} \approx 1 - exp(-\epsilon n/2)$. In the worst case, where all agents happen to value the good at $1$. The optimal revenue would be $OPT(\bar{t}) = n$ whereas if agents apply this unique dominant strategy this mechanism will extract only (a bit over) $n/2$.*

# B Additional Proofs

**Proof**: Given the choice of parameters $\epsilon$ and $q$ truthfulness is established by Lemma **??**.

Let $OPT(t) = \max_{s\in S} F(t, s)$. We need to show that for any type vector $t$ and for large enough $n$, $E_{\hat{M}(t)}(F(t, s)) \geq \max_s F(t, s) - 6 \cdot \sqrt{\frac{d}{\gamma n}}\sqrt{\ln\left(\frac{n\gamma}{2d}|S|\right)}$.

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

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

Applying Theorem 3 we get:

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

Substituting $\epsilon$ for $\sqrt{\frac{\gamma d}{n}}\sqrt{\ln\left(\frac{n\gamma}{2d}|S|\right)}$ we get that for large enough $n$, $\ln\left(\frac{n\epsilon}{2d}|S|\right) \leq \ln\left(\frac{n\gamma}{2d}|S|\right)$. Hence,

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

**QED**