
MECHANISM DESIGN WITH MULTI-ARMED BANDIT

Anonymous authors

Paper under double-blind review

ABSTRACT

A popular approach of automated mechanism design is to formulate a linear program (LP) whose solution gives a mechanism with desired properties. We analytically derive a class of optimal solutions for such an LP that gives mechanisms achieving standard properties of efficiency, incentive compatibility, strong budget balance (SBB), and individual rationality (IR), where SBB and IR are satisfied in expectation. Notably, our solutions are represented by an exponentially smaller number of essential variables than the original variables of LP. Our solutions, however, involve a term whose exact evaluation requires solving a certain optimization problem exponentially many times as the number of players, N , grows. We thus evaluate this term by modeling it as the problem of estimating the mean reward of the best arm in multi-armed bandit (MAB), propose a Probably and Approximately Correct estimator, and prove its asymptotic optimality by establishing a lower bound on its sample complexity. This MAB approach reduces the number of times the optimization problem is solved from exponential to $O(N \log N)$. Numerical experiments show that the proposed approach finds mechanisms that are guaranteed to achieve desired properties with high probability for environments with up to 128 players, which substantially improves upon the prior work.

1 INTRODUCTION

Multi-agent systems can be made efficient by mediators who make system-wide (social) decisions in a way that maximizes social welfare. For example, in a trading network where firms sell goods to each other or to external markets (Hatfield et al., 2013), the mediator can ensure that those traded goods are produced by the firms with the lowest costs and purchased by those with the highest needs (Osogami et al., 2023). While such mediators could maximize their own profit by charging participants (e.g., today’s tech giants who operate consumer marketplaces or ads platforms), they would typically take most of the profit, leaving the participants with only a small portion.

We instead envision an open platform whose purpose is to provide maximal benefits to the participants in multi-agent systems. This is similar to the purpose of designing an auction mechanism for *public* resources, which should be given to those who need them most, and there should be neither budget deficits nor surpluses on the mediator (Bailey, 1997; Cavallo, 2006; Dufton et al., 2021; Gujar & Narahari, 2011; Guo, 2012; Guo & Conitzer, 2009; Manisha et al., 2018; Tacchetti et al., 2022). However, such mechanisms exploit the particular structure of single-sided auctions, where all participants are buyers, and it may be impossible to achieve desired properties in other multi-agent systems such as double-sided auctions (Hobbs et al., 2000; Zou, 2009; Widmer & Leukel, 2016; Stößer et al., 2010; Kumar et al., 2018; Chichin et al., 2017), matching markets (Zhang & Zhu, 2020), and trading networks (Osogami et al., 2023; Wasserkrug & Osogami, 2023) even if those properties could be guaranteed for single-sided auctions.

Here, we design mechanisms for general environments that include all these multi-agent systems, with the primary objectives of efficiency and strong budget balance (SBB). Specifically, we require that the mediator chooses a social decision such that its total value to the players (or participants) is maximized (decision efficiency; DE) and that the expected revenue of the mediator is equal to a target value, $\rho \in \mathbb{R}$ (SBB when $\rho = 0$). As is standard in Bayesian mechanism design (Shoham & Leyton-Brown, 2009), we assume that the value of each social decision to a player depends on the player’s type that is only known to that player, but a joint probability distribution of the players’ types is a common knowledge. Since it would be difficult to achieve DE without knowing the types, we also require that the optimal strategy of each player is to truthfully declare its type regardless of the

054 strategies of the others (dominant strategy incentive compatibility; DSIC). To promote participation,
055 we also require that the expected utility of each player is no less than a target value, $\theta(t_n) \in \mathbb{R}$, that
056 can depend on the type t_n of the player (individual rationality or IR when $\theta \equiv 0$).

057 Although these properties are standard in the literature of mechanism design (Shoham & Leyton-
058 Brown, 2009), here we have introduced parameters, ρ and θ , to generalize the standard definitions
059 of SBB ($\rho = 0$) and IR ($\theta \equiv 0$) with three motivations. First, it is not always possible to achieve
060 the four desired properties with the standard definitions of IR and SBB (Green & Laffont, 1977;
061 Myerson & Satterthwaite, 1983; Osogami et al., 2023), while the generalization will enable the
062 exact characterization of when those properties can be satisfied. Second, this generalization will
063 allow us to develop a principled approach of *learning* a mechanism, where some of the quantities are
064 estimated from samples, *with theoretical guarantees*. The mechanism designed with our learning
065 approach can be shown to satisfy the four desired properties with high probability. Finally, the
066 additional parameters will allow us to model practical requirements. For example, the mediator
067 might need positive revenue to cover the cost of maintaining the platform or might want to guarantee
068 positive expected utility to players to encourage participation in this platform rather than others.

069 We require SBB and IR *in expectation* (*ex ante* or interim) with respect to the distribution of types,
070 while DE and DSIC are satisfied for any realization of types (*ex post*). While these assumptions
071 are similar to those in Osogami et al. (2023); Wasserkrug & Osogami (2023) for trading networks,
072 the *in expectation* properties are certainly weaker than the *ex post* properties usually assumed for
073 auctions (Bailey, 1997; Cavallo, 2006; Dufton et al., 2021; Gujar & Narahari, 2011; Guo, 2012;
074 Guo & Conitzer, 2009; Manisha et al., 2018; Tacchetti et al., 2022). With the weaker properties,
075 however, we derive *analytical* solutions of the mechanisms that satisfy all the four desired properties
076 in the general environments (and characterize when those properties can be satisfied). This is in stark
077 contrast to the prior work, where mechanisms are analytically derived only for auctions with a single
078 type of goods (Bailey, 1997; Cavallo, 2006; Guo, 2011; Guo & Conitzer, 2007; 2009; Moulin, 2009)
079 or with unit demand (Gujar & Narahari, 2011; Guo, 2012). For more complex auctions (Dufton
080 et al., 2021; Manisha et al., 2018; Tacchetti et al., 2022) or trading networks (Osogami et al., 2023;
081 Wasserkrug & Osogami, 2023), mechanisms are computed by numerically solving optimization
082 problems, whose size often grows exponentially with the number of players.

083 While the numerical approaches proposed in Osogami et al. (2023) have been applied only to the
084 trading networks with two players, our analytical solutions can be evaluated numerically with ~ 10
085 players, depending on the number of types. The key bottleneck in our analytical solutions lies in the
086 evaluation of the minimum expected value over possible types of each player. Exact evaluation of
087 this quantity would require computing an efficient social decision for all the K^N combinations of
088 types, where K is the number of possible types of each player, and N is the number of players.

089 To overcome this bottleneck, we model the problem of evaluating this minimum expected value in
090 a multi-armed bandit (MAB) approach (Lattimore & Szepesvári, 2020) and propose an asymptoti-
091 cally optimal learning algorithm for this problem. While the standard objectives of MAB are regret
092 minimization (Auer et al., 1995; 2002) and best arm identification (Audibert et al., 2010; Maron &
093 Moore, 1993; Mnih et al., 2008; Bubeck et al., 2009), our objective is to estimate the mean reward of
094 the best arm. We propose a probably approximately correct (PAC) algorithm, which approximately
095 (with error at most ε) estimates the best mean with high probability (at least $1 - \delta$), and proves that
096 its sample complexity, $O((K/\varepsilon^2) \log(1/\delta))$, matches the lower bound that we derive. This learn-
097 ing approach substantially reduces the number of computing efficient social decisions from K^N to
 $O(K N \log N)$, enabling us to numerically find mechanisms for ~ 100 players, depending on K .

098 Our contributions thus revolve around the optimization problem whose solution gives the mechanism
099 that satisfies DE, SBB, DSIC, and IR (see Section 3-4). First, we establish a sufficient condition that
100 ensures the optimization problem has feasible solutions, and prove that this sufficient condition is
101 also necessary when the players have independent types (see Section 5). Second, for cases where
102 this sufficient condition holds, we analytically derive a class of optimal solutions to this optimization
103 problem, which in turn gives mechanisms that satisfy DE, SBB, DSIC, and IR for general environ-
104 ments including auctions and trading networks (see Section 5). Third, we model the problem of
105 evaluating a quantity in the above analytical expressions as best mean estimation in MAB, propose
106 a PAC algorithm for this problem, and prove its asymptotic optimality (see Section 6). Finally, we
107 empirically validate the effectiveness of the proposed approach (see Section 7). In Section 2, we
start by positioning our contributions to the prior work.

2 RELATED WORK

The prior work most related to ours is Osogami et al. (2023), which formulates and numerically solves the LP whose solution gives the mechanism for a trading network that satisfies DE, DSIC, IR, and weak budget balance (WBB; expected revenue of the mediator is nonnegative). While the objective of the LP is rather arbitrary and SBB is given just as an example in Osogami et al. (2023), we focus on SBB and drive analytical solutions for this particular objective. Our formulation extends Osogami et al. (2023) with additional parameters and notations that cover environments beyond trading networks, but these extensions are relatively straightforward.

In the rest of this section, we discuss related work on mechanism design, with a focus on those aiming to achieve SBB, as well as related work on MAB with a focus on PAC algorithms. In particular, we will see that our learning approach is unique in that we estimate a particular quantity in the optimal solution that we derive *analytically*, and this leads us to propose a new PAC algorithm and establish its optimality for an underexplored objective of best mean estimation in MAB.

In single-sided auctions where only buyers make strategic decisions, Vickrey–Clarke–Groves (VCG) mechanisms with Clark’s pivot rule (also called VCG auction) satisfy *ex post* DE (called allocative efficiency in auctions), DSIC, IR, and WBB (Nisan, 2007). However, the Green-Laffont Impossibility Theorem implies that no mechanism can guarantee DE, DSIC, and SBB simultaneously for all environments (Green & Laffont, 1977; 1979). This has led to a series of work on redistributing the revenue of the mediator to the players as much as possible (i.e., to make budget balance as strong as possible), while satisfying DSIC, DE, IR, and WBB. For auctions with single or homogeneous goods (Bailey, 1997; Cavallo, 2006; Guo, 2011; Guo & Conitzer, 2007; 2009; Moulin, 2009) or for auctions where players have unit demand (Gujar & Narahari, 2011; Guo, 2012), researchers have derived analytical solutions that optimally redistribute the payment to the players. For auctions with multi-unit demands on heterogeneous goods, the prior work has proposed numerical approaches that seek to find the piecewise linear functions (Dufton et al., 2021) or neural networks (Manisha et al., 2018; Tacchetti et al., 2022) that best approximate the optimal redistribution functions.

We consider the environments that not only allow heterogeneous goods and multi-unit demands but also are more general than single-sided auctions. In particular, our players may have negative valuation on a social decision. The Myerson-Satterthwaite Impossibility Theorem (Myerson & Satterthwaite, 1983) thus implies that, unlike VCG auctions, no mechanism can guarantee *ex post* DE, DSIC, IR, and WBB simultaneously for all the environments that we consider. We thus derive mechanisms that achieve DE, DSIC, IR, and SBB in the best possible manner. A limitation of our results is that IR and SBB are satisfied only in expectation. Such a guarantee in expectation can however be justified for risk-neutral mediator and players (Osogami et al., 2023). Our model can also guarantee *strictly* positive expected utility, which in turn can ensure nonnegative utility with high probability when the player repeatedly participate in the mechanism.

For auctions, there also exists a large body of the literature on maximizing the revenue of the mediator (Myerson, 1981) with recent focus on automated mechanism design (AMD) with machine learning (Duetting et al., 2019; Rahme et al., 2021; Ivanov et al., 2022; Curry et al., 2020) and analysis of its sample complexity (Balcan et al., 2016; Morgenstern & Roughgarden, 2015; Syrgkanis, 2017). Similar to these and other studies of AMD (Sandholm, 2003; Conitzer & Sandholm, 2002), we formulate an optimization problem whose solution gives the mechanism with desired properties. However, instead of solving it numerically, we analytically derive optimal solutions. Also, while the prior work analyzes the sample complexity for finding the mechanism that maximizes the expected revenue, we analytically find optimal mechanisms and characterize the sample complexity of evaluating a particular expression in the analytically designed mechanisms.

We evaluate our analytical expression through best mean estimation (BME) in MAB, where the standard objectives are regret minimization (Auer et al., 1995; 2002) and best arm identification (BAI) (Audibert et al., 2010; Maron & Moore, 1993; Mnih et al., 2008; Bubeck et al., 2009). The prior work on MAB that is most relevant to ours is PAC learning for BAI and analysis of its sample complexity. We reduce the problem of BME to BAI and prove the lower bound on the sample complexity of BME using a technique known for BAI (Even-Dar et al., 2002). However, while this technique does not give tight lower bound for BAI (Mannor & Tsitsiklis, 2004), we show that it gives tight lower bound for BME. Notice that the problem of estimating the best mean frequently appears in reinforcement learning (van Hasselt, 2010) and machine learning (Kajino et al., 2023),

but there the focus is on how best to estimate the best mean with a given set of samples (van Hasselt, 2013), while our focus is on how best to collect samples to estimate the best mean.

3 SETTINGS

The goal of mechanism design is to specify the rules of a game in a way that an outcome desired by the mechanism designer is achieved when rational players (i.e., players whose goal it is to maximize their individual utility) participate in that game (Jackson, 2014; Shoham & Leyton-Brown, 2009). Formally, let $\mathcal{N} := [1, N]$ be the set of players and \mathcal{O} be the set of possible outcomes. For each player $n \in \mathcal{N}$, let \mathcal{A}_n be the set of available actions and \mathcal{T}_n be the set of possible types. Let $\mathcal{A} := \mathcal{A}_1 \times \dots \times \mathcal{A}_N$ and $\mathcal{T} := \mathcal{T}_1 \times \dots \times \mathcal{T}_N$ be the corresponding product spaces. A mechanism $\mu : \mathcal{A} \rightarrow \mathcal{O}$ determines an outcome depending of the actions taken by the players. Let $u_n : \mathcal{O} \times \mathcal{T}_n \rightarrow \mathbb{R}$ be the utility function of each player $n \in \mathcal{N}$.

We consider Bayesian games where the players' types follow a probability distribution that is known to all players and the mediator. Before selecting actions, the players know their own types but not the types of the other players. A strategy of each player $n \in \mathcal{N}$ is thus a function from \mathcal{T}_n to \mathcal{A}_n .

We assume that an outcome is determined by a social decision and payment; hence, a mechanism μ consists of a decision rule and a payment rule. Let \mathcal{D} be the set of possible social decisions. Given the actions of the players, the decision rule $\phi : \mathcal{A} \rightarrow \mathcal{D}$ determines a social decision, and the payment rule $\tau : \mathcal{A} \rightarrow \mathbb{R}^{\mathcal{N}}$ determines the amount of (possibly negative) payment to the mediator from each player. Let $v : \mathcal{D} \times (\mathcal{T}_1 \cup \dots \cup \mathcal{T}_N) \rightarrow \mathbb{R}$ specify the value of a given social decision to the player of a given type. Then the utility of player i when players take actions $a \in \mathcal{A}$ is

$$u_n(\mu(a); t_n) = u_n((\phi(a), \tau(a)); t_n) = v(\phi(a); t_n) - \tau_n(a). \quad (1)$$

Throughout, we assume that \mathcal{N} , \mathcal{D} , and $\mathcal{T}_n, \forall n \in \mathcal{N}$ are finite sets.

Without loss of generality by the revelation principle (Shoham & Leyton-Brown, 2009), we consider only direct mechanisms, where the action available to each player is to declare which type the player belongs to from the set of possible types (i.e., $\mathcal{A}_n = \mathcal{T}_n, \forall n \in \mathcal{N}$). We will thus use \mathcal{T}_n for \mathcal{A}_n .

Then we seek to achieve the following four properties with our mechanisms. The first property is Dominant Strategy Incentive Compatibility (DSIC), which ensures that the optimal strategy of each player is to truthfully reveal its type regardless of the strategies of the other players. Formally,

$$[\text{DSIC}] \quad v(\phi(t_n, t'_{-n}); t_n) - \tau_n(t_n, t'_{-n}) \geq v(\phi(t'; t_n) - \tau_n(t'), \forall t' \in \mathcal{T}, \forall t_n \in \mathcal{T}_n, \forall n \in \mathcal{N}, \quad (2)$$

where the left-hand side represents the utility of the player having type t_n when it declares the same t_n , and the other players declare arbitrary types t'_{-n} .

The second property is Decision Efficiency (DE), which requires that the mediator chooses the social decision that maximizes the total value to the players. With DSIC, we can assume that the players declare true types, and hence we can write DE as a condition on the decision rule:

$$[\text{DE}] \quad \phi(t) \in \operatorname{argmax}_{d \in \mathcal{D}} \sum_{n \in \mathcal{N}} v(d; t_n) \quad \forall t \in \mathcal{T}. \quad (3)$$

As the third property, we generalize individual rationality and require that the expected utility of each player is at least as large as a target value that can depend on its type. We refer to this property as θ -IR. Again, assuming that players declare true types due to DSIC, we can write θ -IR as follows:

$$[\theta\text{-IR}] \quad \mathbb{E}[v(\phi(t); t_n) - \tau_n(t) \mid t_n] \geq \theta(t_n) \quad \forall t_n \in \mathcal{T}_n, \forall n \in \mathcal{N}, \quad (4)$$

where $\theta : \mathcal{T}_1 \cup \dots \cup \mathcal{T}_N \rightarrow \mathbb{R}$ determines the target expected utility for each type. Throughout (except in Section 6, where we discuss general MAB models), \mathbb{E} denotes the expectation with respect to the probability distribution \mathbb{P} of types, which is the only probability that appears in our mechanisms.

The last property is a generalization of Budget Balance (BB), which we refer to as ρ -WBB and ρ -SBB. Specifically, ρ -WBB requires that the expected revenue of the mediator is no less than a given constant $\rho \in \mathbb{R}$, and ρ -SBB requires that it is equal to ρ . Again, assuming that the players declare true types due to DSIC, these properties can be written as follows:

$$[\rho\text{-WBB}] \quad \sum_{n \in \mathcal{N}} \mathbb{E}[\tau_n(t)] \geq \rho. \quad [\rho\text{-SBB}] \quad \sum_{n \in \mathcal{N}} \mathbb{E}[\tau_n(t)] = \rho. \quad (5)$$

While ρ -SBB is stronger than ρ -WBB, we will see that ρ -SBB is satisfiable iff ρ -WBB is satisfiable.

4 OPTIMIZATION PROBLEM FOR AUTOMATED MECHANISM DESIGN

Following Osogami et al. (2023), we seek to find optimal mechanisms in the class of VCG mechanisms. A VCG mechanism is specified by a pair (ϕ^*, h) . Specifically, after letting player take the actions of declaring their types $t \in \mathcal{T}$, the mechanism first finds a social decision $\phi^*(t)$ using a decision rule $\phi = \phi^*$ that satisfies DE (3). It then determines the amount of payment from each player $n \in \mathcal{N}$ to the mediator by

$$\tau_n(t) = h_n(t_{-n}) - \sum_{m \in \mathcal{N}_{-n}} v(\phi^*(t); t_m), \quad (6)$$

where we define $\mathcal{N}_{-n} := \mathcal{N} \setminus \{n\}$, and $h_n : \mathcal{T}_{-n} \rightarrow \mathbb{R}$ is an arbitrary function of the types of the players other than n and referred to as a pivot rule. The decision rule ϕ^* guarantees DE (3) by construction, and the payment rule (6) then guarantees DSIC (2) (Nisan, 2007).

Our problem is now reduced to find the pivot rule, $h = \{h_n\}_{n \in \mathcal{N}}$, that minimizes the expected revenue of the mediator, while satisfying θ -IR and ρ -WBB. This may lead to satisfying ρ -SBB if the revenue is maximally reduced. To represent this reduced problem, let

$$w^*(t) := \sum_{n \in \mathcal{N}} v(\phi^*(t); t_n) \quad (7)$$

be the total value of the efficient social decision when the players have types t . Then we can rewrite θ -IR (for the player having type t_n) and ρ -WBB as follows (see Appendix A.1 for full derivation):

$$\mathbb{E}[v(\phi^*(t); t_n) - \tau_n(t) \mid t_n] \geq \theta(t_n) \iff \mathbb{E}[w^*(t) \mid t_n] - \mathbb{E}[h_n(t_{-n}) \mid t_n] \geq \theta(t_n) \quad (8)$$

$$\sum_{n \in \mathcal{N}} \mathbb{E}[\tau_n(t)] \geq \rho \iff \sum_{n \in \mathcal{N}} \mathbb{E}[h_n(t_{-n})] - (N-1) \mathbb{E}[w^*(t)] \geq \rho. \quad (9)$$

Therefore, we arrive at the following linear program (LP):

$$\min_h \quad \sum_{n \in \mathcal{N}} \mathbb{E}[h_n(t_{-n})] \quad (10)$$

$$\text{s.t.} \quad \mathbb{E}[w^*(t) \mid t_n] - \mathbb{E}[h_n(t_{-n}) \mid t_n] \geq \theta(t_n) \quad \forall t_n \in \mathcal{T}_n, \forall n \in \mathcal{N} \quad (11)$$

$$\sum_{n \in \mathcal{N}} \mathbb{E}[h_n(t_{-n})] - (N-1) \mathbb{E}[w^*(t)] \geq \rho. \quad (12)$$

The approach of Osogami et al. (2023) is to numerically solve this LP possibly with approximations. Since $h_n(t_{-n})$ is a variable for each $t_{-n} \in \mathcal{T}_{-n}$ and each $n \in \mathcal{N}$, the LP has $N K^{N-1}$ variables and $N K + 1$ constraints, when each player has K possible types. When this LP is feasible, let h^* be its optimal solution; then the VCG mechanism (ϕ^*, h^*) guarantees DSIC, DE, θ -IR, and ρ -WBB (formalized as Proposition 4 in Appendix A.1). Otherwise, no VCG mechanisms can guarantee them all. In the next section, we characterize exactly when the LP is feasible and provide analytical solutions to the LP.

5 ANALYTICAL SOLUTION TO THE OPTIMIZATION PROBLEM

We first establish a sufficient condition and a necessary condition for the LP to have feasible solutions. Note that complete proofs for all theoretical statements are provided in Appendix B.

Lemma 1. *The LP (10)-(12) is feasible if*

$$\sum_{n \in \mathcal{N}} \min_{t_n \in \mathcal{T}_n} \{\mathbb{E}[w^*(t) \mid t_n] - \theta(t_n)\} \geq (N-1) \mathbb{E}[w^*(t)] + \rho. \quad (13)$$

Lemma 2. *When types are independent (t_m and t_n are independent for any $m \neq n$ under \mathbb{P}), the LP (10)-(12) is feasible only if (13) holds.*

These two lemmas establish the following necessary and sufficient condition.

Corollary 1. *When types are independent, the LP (10)-(12) is feasible if and only if (13) holds.*

While the LP is not necessarily feasible, one may choose θ and ρ in a way that it is guaranteed to be feasible. For example, the feasibility is guaranteed (i.e., (13) is satisfied) by setting

$$\theta \equiv 0 \quad \text{and} \quad \rho = \left[\sum_{n \in \mathcal{N}} \min_{t_n \in \mathcal{T}_n} \mathbb{E}[w^*(t) | t_n] - (N-1) \mathbb{E}[w^*(t)] \right]^-, \quad (14)$$

where $[x]^- := \min\{x, 0\}$ for $x \in \mathbb{R}$. When $\rho < 0$, the mediator might get negative expected revenue, but the expected loss of the mediator is at most $|\rho|$. Appendix A.2 provides alternative ρ and θ that guarantee feasibility, but some of the players might incur negative expected utility.

Finally, we derive a class of optimal solutions to the LP when it is feasible.

Lemma 3. *A pivot rule is said to be constant if and only if, for each $n \in \mathcal{N}$, there exists a constant η_n such that $h_n(t_{-n}) = \eta_n, \forall t_{-n} \in \mathcal{T}_{-n}$. Let \mathcal{H} be the set of constant pivot rules that satisfy:*

$$\eta_n = \min_{t_n \in \mathcal{T}_n} \{\mathbb{E}[w^*(t) | t_n] - \theta(t_n)\} - \delta_n \quad \forall n \in \mathcal{N} \quad (15)$$

where δ lies on the following simplex:

$$\delta_n \geq 0 \quad \forall n \in \mathcal{N} \quad (16)$$

$$\sum_{n \in \mathcal{N}} \delta_n = \sum_{n \in \mathcal{N}} \min_{t_n \in \mathcal{T}_n} \{\mathbb{E}[w^*(t) | t_n] - \theta(t_n)\} - (N-1) \mathbb{E}[w^*(t)] - \rho. \quad (17)$$

Then, when (13) holds, \mathcal{H} is nonempty, and any $h \in \mathcal{H}$ is an optimal solution to the LP (10)-(12).

When types are independent, the condition (13) is necessary for the existence of a feasible solution; hence, we do not lose optimality by considering only the solutions in \mathcal{H} . On the other hand, when types are dependent, the condition (13) may still be satisfied, where the solutions in \mathcal{H} remain optimal. However, dependent types no longer necessitate (13) to satisfy (11)-(12); hence the space of constant pivot rules may not suffice to find an optimal solution if (13) does not hold.

Proposition 1. *The LP (10)-(12) may be feasible even if (13) is violated (when types are dependent).*

Corollary 2. *Feasible solutions to the LP (10)-(12) include at least one constant pivot rule (defined in Lemma 3) if and only if (13) holds (whether types are dependent or not).*

In deriving the optimal solutions, we have substantially reduced the essential number of variables (from NT^{N-1} to N when each player has T types). Our approach can therefore not only find but also represent or store solutions exponentially more efficiently than Osogami et al. (2023).

It turns out that the solutions in \mathcal{H} not only satisfy ρ -WBB but also ρ -SBB (in addition to DE, DSIC, and θ -IR) regardless of whether the types are independent or not:

Corollary 3. *Any VCG mechanism given with a pivot rule in \mathcal{H} satisfies ρ -SBB.*

When the LP is infeasible, we can construct a mechanism that satisfies one of ρ -SBB and θ -IR (in addition to DE and DSIC) regardless of whether the types are independent or not:

Corollary 4. *Any VCG mechanism with a pivot rule given by (15) and (17) satisfies ρ -SBB.*

Corollary 5. *Any VCG mechanism with a pivot rule given by (15) and (16) satisfies θ -IR for any $t_n \in \mathcal{T}_n$ and $n \in \mathcal{N}$.*

For example, for any $t_n \in \mathcal{T}_n$ and $n \in \mathcal{N}$, the following pivot rule always satisfies θ -IR:

$$\eta_n = \min_{t_n \in \mathcal{T}_n} \{\mathbb{E}[w^*(t) | t_n] - \theta(t_n)\} - \max\{\delta, 0\}, \quad (18)$$

where we define (see also (58) in the appendix)

$$\delta := \frac{1}{N} \left(\sum_{n \in \mathcal{N}} \min_{t_n \in \mathcal{T}_n} \{\mathbb{E}[w^*(t) | t_n] - \theta(t_n)\} - (N-1) \mathbb{E}[w^*(t)] - \rho \right). \quad (19)$$

So far, we have analytically derived the class of optimal solutions for the LP. Although these analytical solutions substantially reduce the computational cost of solving the LP compared to the numerical solutions in Osogami et al. (2023), they still need to evaluate

$$\min_{t_n \in \mathcal{T}_n} \{\mathbb{E}[w^*(t) | t_n] - \theta(t_n)\} \quad (20)$$

for each $n \in \mathcal{N}$. Since \mathbb{E} is the expectation with respect to the distribution \mathbb{P} over \mathcal{T} , this would require evaluating $w^*(t)$ for all $t \in \mathcal{T}$. Recall that $w^*(t)$ defined with (7) is the total value of the efficient decision $\phi^*(t)$, which is given by a solution to an optimization problem (3). Without any structure in \mathcal{D} and v , this would require evaluating the total value for all decisions in \mathcal{D} .

6 ONLINE LEARNING FOR EVALUATING THE ANALYTICAL SOLUTION

To alleviate the computational complexity associated with (20), we propose a learning approach. A key observation is that the problem of estimating (20) can be considered as a variant of a multi-armed bandit (MAB) problem whose objective is to estimate the mean reward of the best arm. Specifically, the MAB gives the reward $\theta(t_n) - w^*(t)$ when we pull an arm $t_n \in \mathcal{T}_n$, where t is a sample from the conditional distribution $\mathbb{P}[\cdot | t_n]$. Following the relevant prior work on MAB (Even-Dar et al., 2002; 2006; Hassidim et al., 2020; Mannor & Tsitsiklis, 2004), we maximize reward in this section.

Since we assume that \mathcal{N} , \mathcal{D} , and $\mathcal{T}_n, \forall n \in \mathcal{N}$ are finite, there exist constants, $\bar{\theta}$ and \bar{v} , such that $|\theta(t')| \leq \bar{\theta}$ and $|v(d; t')| \leq \bar{v}, \forall d \in \mathcal{D}, \forall t' \in \cup_{n \in \mathcal{N}} \mathcal{T}_n$. Then we can also bound $|\theta(t_n) - w^*(t)| \leq \bar{\theta} + N \bar{v}, \forall t \in \mathcal{T}, \forall n \in \mathcal{N}$. Namely, the reward in the MAB is bounded. We assume that we know the bounds on the reward and can scale it such that the scaled reward is in $[0, 1]$ surely.

We also assume that we have access to an arbitrary size of the sample that is independent and identically distributed (i.i.d.) according to $\mathbb{P}[\cdot | t_n]$ for any $t_n \in \mathcal{T}_n, n \in \mathcal{N}$. When players have independent types, such sample can be easily constructed as long as we have access to i.i.d. sample $\{t^{(i)}\}_{i=1,2,\dots}$ from \mathcal{T} , because $\{(t_n, t_{-n}^{(i)})\}_i$ is the sample from $\mathbb{P}[\cdot | t_n]$ for any $t_n \in \mathcal{T}_n, n \in \mathcal{N}$.

Consider the general K -armed bandit where the reward of each arm is bounded in $[0, 1]$. For each $k \in [1, K]$, let μ_k be the true mean of arm k . Let $\mu_* := \max_k \mu_k$ be the best mean-reward, which we want to estimate. We say that the sample complexity of an algorithm for a MAB is T if the sample size needed by the algorithm is bounded by T (i.e., arms are pulled at most T times).

A standard Probably Approximately Correct (PAC) algorithm for MAB returns an ε -optimal arm with probability at least $1 - \delta$ for given ε, δ (Even-Dar et al., 2006; Hassidim et al., 2020). We will use the following definition:

Definition 1 ((ε, δ) -PAC Best Arm Identifier (Even-Dar et al., 2006)). *For $\varepsilon, \delta > 0$, we say that an algorithm is (ε, δ) -PAC Best Arm Identifier (BAI) if the arm \hat{I} identified by the algorithm satisfies*

$$\Pr(\mu_{\hat{I}} \geq \mu_* - \varepsilon) \geq 1 - \delta. \quad (21)$$

Definition 1 is different from what we need to evaluate (20). Formally, we need

Definition 2 ((ε, δ) -PAC Best Mean Estimator). *For $\varepsilon, \delta > 0$, we say that an algorithm is (ε, δ) -PAC Best Mean Estimator (BME) if the best mean $\hat{\mu}$ estimated by the algorithm satisfies*

$$\Pr(|\hat{\mu} - \mu_*| \leq \varepsilon) \geq 1 - \delta. \quad (22)$$

BAI and BME are related but different. For example, consider the case where the best arm has large variance and $\mu_* = 1/2$, and the other arms always give zero reward $\mu_n = 0, \forall n \neq *$. Then relatively small sample would suffice for BAI due to the large gap $\mu_* - \mu_n = 1/2, \forall n \neq *$, while BME would require relatively large sample due to the large variance of the best arm $*$. As another example, consider the case where we have many arms whose rewards follow Bernoulli distributions. Suppose that half of the arms have an expected value of 1, and the other half have an expected value of $1 - (3/2)\varepsilon$. Then by pulling sufficiently many arms (once for each arm), we can estimate that the best mean is at least $1 - \varepsilon$ with high probability (by Hoeffding's inequality), but BAI would require pulling arms sufficiently many times until we pull one of the best arms sufficiently many times (to be able to say that this particular arm has an expected value at least $1 - \varepsilon$).

For (ε, δ) -PAC BAI, the following lower and upper bounds on the sample complexity are known:

Proposition 2 (Mannor & Tsitsiklis (2004)). *There exists a (ε, δ) -PAC BAI with sample complexity $O((K/\varepsilon^2) \log(1/\delta))$, and any (ε, δ) -PAC BAI must have the sample complexity at least $\Omega((K/\varepsilon^2) \log(1/\delta))$.*

We establish the same lower and upper bounds for (ε, δ) -PAC BME:

Theorem 1. *There exists a (ε, δ) -PAC BME with sample complexity $O((K/\varepsilon^2) \log(1/\delta))$, and any (ε, δ) -PAC BME must have the sample complexity at least $\Omega((K/\varepsilon^2) \log(1/\delta))$.*

Our upper bound is established by reducing BME to BAI. Suppose that we have access to an arbitrary (ε, δ) -PAC BAI with sample complexity M (Algorithm 2 in Appendix A.3). We can construct a $(\frac{3}{2}\varepsilon, 2\delta)$ -PAC BME by first running the (ε, δ) -PAC BAI and then sampling from the arm \hat{I} that is identified as the best until we have a sufficient number, m^* , of samples from \hat{I} (Algorithm 3 in Appendix A.3). When m^* is appropriately selected, the following lemma holds:

Lemma 4. *When a (ε, δ) -PAC BAI with sample complexity M is used, Algorithm 3 has sample complexity $M + m^*$, where $m^* := \lceil (2/\varepsilon^2) \log(1.22/\delta) \rceil$, and returns $\hat{\mu}_{\hat{I}}$ that satisfies*

$$\Pr \left(|\hat{\mu}_{\hat{I}} - \mu_{\star}| > \frac{3}{2}\varepsilon \right) \leq 2\delta. \quad (23)$$

By Proposition 2, this establishes the upper bound on the sample complexity in Theorem 1.

Our lower bound is established by showing that an arbitrary (ε, δ) -PAC BME can be used to solve the problem of identifying whether a given coin is negatively or positively biased (precisely, ε -Biased Coin Problem of Definition 3 in Appendix A.3), for which any algorithm is known to require *expected* sample complexity at least $\Omega((1/\varepsilon^2) \log(1/\delta))$ to solve correctly with probability at least $1 - \delta$ (Chernoff, 1972; Even-Dar et al., 2002) (see Lemma 7 in Appendix A.3). The following lemma, together with Lemma 7, establishes the lower bound on the sample complexity in Theorem 1:

Lemma 5. *If there exists an $(\varepsilon/2, \delta/2)$ -PAC BME with sample complexity M for K -armed bandit, then there also exists an algorithm, having expected sample complexity M/K , that can solve the ε -Biased Coin Problem correctly with probability at least $1 - \delta$.*

We prove Lemma 5 by reducing the ε -Biased Coin Problem to BME. This proof technique was also used in Even-Dar et al. (2002) to prove a lower bound on the sample complexity of BAI. What is interesting, however, is that the lower bound in Even-Dar et al. (2002) is not tight when $\delta < 1/K$, and a tight lower bound is established by a different technique in Mannor & Tsitsiklis (2004). On the other hand, our proof gives a tight lower bound on the sample complexity of BME. In Appendix A.3, we further discuss where this difference in the derived lower bounds comes from.

Finally, we connect the results on BME back to our mechanism in Section 5 and provide a guarantee on the properties of the mechanism when the term (20) is estimated with BME. To this end, we define the terms involving expectations in Lemma 3 as follows:

$$\kappa_n(\theta) := \min_{t_n \in \mathcal{T}_N} \{ \mathbb{E}[w^* | t_n] - \theta(t_n) \} \quad (24)$$

$$\lambda(\rho) := \mathbb{E}[w^*(t)] + \rho/(N - 1). \quad (25)$$

Then, recalling that N is the number of players, we have the following lemma:

Lemma 6. *Let $\tilde{\kappa}_n(\theta)$ for $n \in \mathcal{N}$ and $\tilde{\lambda}(\rho)$ be independent estimates of $\kappa_n(\theta)$ and $\lambda(\rho)$ respectively given by an (ε', δ') -PAC Best Mean Estimator and a standard (ε'', δ') -PAC estimator of expectation. Also, let $\tilde{d} := d(\tilde{\kappa}(\theta), \tilde{\lambda}(\rho), \varepsilon''', \varepsilon''')$ be a point on the following simplex (here, we change the notation from δ in Lemma 3 to \tilde{d} to avoid confusion):*

$$\tilde{d}_n \geq \varepsilon''', \forall n \in \mathcal{N} \quad (26)$$

$$\sum_{n \in \mathcal{N}} \tilde{d}_n = \sum_{n \in \mathcal{N}} \tilde{\kappa}_n(\theta) - (N - 1) (\tilde{\lambda}(\theta) + \varepsilon''') \quad (27)$$

Then the VCG mechanism with the constant pivot rule $h_n(t_{-n}) = \eta_n = \tilde{\kappa}_n(\theta) - \tilde{d}_n$ satisfies $(\theta - (\varepsilon' - \varepsilon'''))$ -IR and $(\rho - (N - 1)(\varepsilon''' - \varepsilon'''))$ -WBB with probability $(1 - \delta')^{N+1}$.

Notice that the sufficient condition of Lemma 1 states that the simplex in Lemma 3 is nonempty. Analogously, when the simplex in Lemma 6 is empty, we cannot provide the solution that guarantees the properties stated in Lemma 6. Since DSIC and DE remain satisfied regardless of whether κ_n for $n \in \mathcal{N}$ and λ are estimated or exactly computed, Lemma 6 immediately establishes the following theorem by appropriately choosing the parameters:

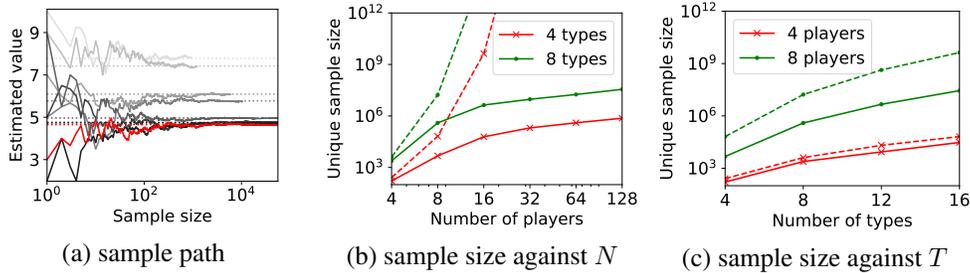


Figure 1: (a) Representative sample path that shows the estimated values of $\mathbb{E}[w^*(t) \mid t_n]$ for $t_n \in \mathcal{T}_n$ against sample size used by a $(0.25, 0.1)$ -PAC BME with Successive Elimination (SE-BME; Algorithm 4) for 8 players, each with 8 types. (b)-(c) The unique sample size (the number of unique t which $w^*(t)$ is evaluated with) required by the exact computation of $\min_{t_n \in \mathcal{T}_n} \mathbb{E}[w^*(t) \mid t_n], \forall n \in \mathcal{N}$ (dashed curves) and by SE-BME (solid curves).

Theorem 2. In Lemma 6, let $\varepsilon''' = \varepsilon'$, $\varepsilon'''' = \varepsilon''$, and $\delta' = 1 - (1 - \delta)^{1/(N+1)}$. Then the VCG mechanism with the constant pivot rule $h_n(t_{-n}) = \tilde{\kappa}_n(\theta) - \tilde{d}_n$ satisfies DSIC, DE, θ -IR, and ρ -WBB with probability $1 - \delta$.

Recall that the exact computation of our mechanism in Lemma 3 requires evaluating $w^*(t)$ for all $t \in \mathcal{T}$, whose size grows exponentially with the number of players, N . Our BME algorithm reduces this requirement to $O(N \log N)$, as is formally proved in the following proposition:

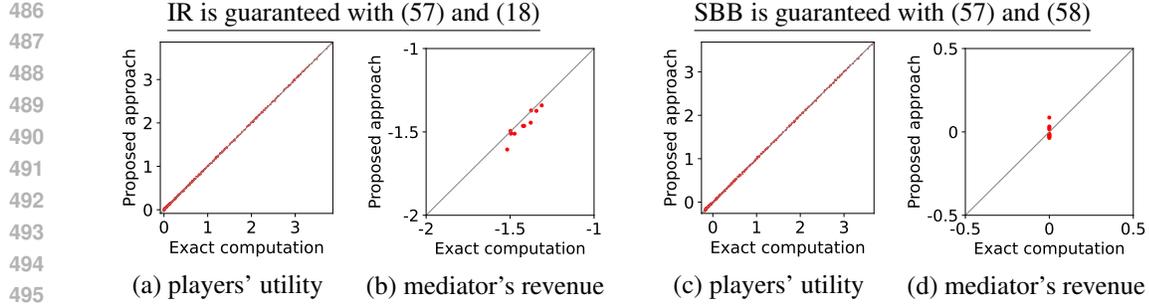
Proposition 3. The sample complexity to learn the constant pivot rule in Theorem 2 is $O((N K/\varepsilon^2) \log(N/\delta))$, where $N = |\mathcal{N}|$ is the number of players, and $K = \max_{n \in \mathcal{N}} |\mathcal{T}_n|$ is the maximum number of types of each player.

7 NUMERICAL EXPERIMENTS

We conduct several numerical experiments to validate the effectiveness of the proposed approach and to understand its limitations. Specifically, we design our experiments to investigate the following questions. i) BME studied in Section 6 has asymptotically optimal sample complexity, but how well can we estimate the best mean when there are only a moderate number of arms? ii) How much can BME reduce the number of times $w^*(t)$ given by (7) is evaluated? iii) How well θ -IR and ρ -SBB are satisfied when (20) is estimated by BME rather than calculated exactly? In this section, we only provide a brief overview of our experiments; for full details, see Appendix A.4. In particular, we find the key empirical property of BME that it can reduce the computational complexity by many orders of magnitude when the number of players is moderate (~ 10) to large but is relatively less effective against the increased number of types.

For question i), we compare Successive Elimination for BAI (SE-BAI), which is known to perform well for a moderate number of arms (Hassidim et al., 2020; Even-Dar et al., 2006), against the corresponding algorithm for BME (SE-BME) and summarize the results in Appendix A.4.1. Overall, we find that SE-BME generally requires smaller sample size than SE-BAI, except when there are only a few arms and their means have large gaps. The efficiency of SE-BAI for this case makes intuitive sense, since the best arm can be identified without estimating the means with high accuracy.

For question ii), we quantitatively validate the effectiveness of SE-BME in reducing the number of times $w^*(t)$ is computed when we evaluate (20) in a setting of mechanism design, as is detailed in Appendix A.4.2. Figure 1 summarizes the results. Panel (a) shows that the means close to the minimum value are estimated with high accuracy, while others are eliminated by SE-BME after a small number of samples. Panels (b)-(c) show that SE-BME (solid curves) evaluates $w^*(t)$ by orders of magnitude smaller number of times than what is required by exact computation (dashed curves). The effectiveness of SE-BME is particularly striking when N (the number of players) grows. While exact computation evaluates $w^*(t)$ exponentially many times as N grows, SE-BME has only linear dependency on N . This insensitivity of the sample complexity of SE-BME to N makes intuitive sense, since N only affects the distribution of the reward and keeps the number of arms unchanged.



497 Figure 2: The red dot shows expected utility in (a) and (c) and expected revenue in (b) and (d), when analytical solutions are evaluated exactly (horizontal) or estimated with $(0.25, 0.1)$ -PAC SE-BME (vertical) for environments with 8 players, each having 8 possible types. The analytical solution guarantees IR in (a) and (b) and SBB in (c) and (d). Results are plotted for 10 random seeds.

503 For question iii), we quantitatively evaluate how well θ -IR and ρ -SBB are guaranteed when we estimate (20) with SE-BME, where we continue to use the same setting of mechanism design. In Figure 2, we study two analytical solutions: one guarantees to satisfy θ -IR (with $\theta \equiv 0$), and the other guarantees to satisfy ρ -SBB (with $\rho = 0$). The red dots show the expected utility of the players in (a) and (c) as well as the expected revenue of the mediator in (b) and (d), when (20) is evaluated either exactly (horizontal axes) or estimated with SE-BME (vertical axes). Overall, the expected revenue of the mediator is more sensitive than the expected utility of the players to the error in the estimation of (20), because (12) involves the summation $\sum_{n \in \mathcal{N}} \eta_n$, while (11) only involves η_n for a single $n \in \mathcal{N}$. We can however ensure that ρ -WBB (instead of SBB) is satisfied with high probability by replacing the ρ with a $\rho' > \rho$ when we compute η_n after (20) is estimated. As we show and explain with Figure 8 in Appendix A.4.3, this will simply shift the dots in the figure, and we can guarantee ρ -WBB with high probability with an appropriate selection of ρ' .

515 We have run all of the experiments on a single core with at most 66 GB memory without GPUs in a cloud environment, as is detailed in Appendix A.4.4. The associated source code is submitted as a supplementary material and will be open-sourced upon acceptance.

520 8 CONCLUSION

522 We have analytically derived optimal solutions for the LP that gives mechanisms that guarantee the desired properties of DE, DSIC, ρ -SBB, and θ -IR. When there are N players, each with T possible types, the LP involves $N T^{N-1}$ variables, while our analytical solutions are represented by only N essential variables. While Osogami et al. (2023) numerically solves this LP only for $N = T = 2$, we have exactly evaluated our analytical solutions for $N = T = 8$ (see Figure 1). The analytical solution, however, involves a term whose exact evaluation requires finding efficient social decisions T^N times. We have modeled the problem of evaluating this term as best mean estimation in multi-armed bandit, proposed a PAC estimator, and proved its asymptotic optimality by establishing a lower bound on the sample complexity. Our experiments show that our PAC estimator enables finding mechanism for $N = 128$ and $T = 8$ with a guarantee on the desired properties.

532 The proposed approach makes a major advancement in the field and can positively impact society by providing guarantees on desired properties for large environments, which existing approaches are unable to handle. However, it presents certain limitations and potential challenges, which inspire several directions for further research. Regarding limitations, one should keep in mind that our mechanisms guarantee ρ -SBB and θ -IR only in expectation, and that the our approaches have only been applied to the environments up to 128 players and up to 16 types in our experiments. Regarding societal impacts, our approach might result in the mechanisms that are unfair to some of the players, because DE does not mean that all of the players are treated fairly. These limitations and societal impacts are further discussed in Appendix A.5.

540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593

REFERENCES

- Jean-Yves Audibert, Sébastien Bubeck, and Rémi Munos. Best arm identification in multi-armed bandits. In Adam Tauman Kalai and Mehryar Mohri (eds.), *Proceedings of the 23rd Conference on Learning Theory*, pp. 41–53. Omnipress, June 2010. URL <http://colt2010.haifa.il.ibm.com/papers/COLT2010proceedings.pdf#page=49>.
- P. Auer, N. Cesa-Bianchi, Y. Freund, and R.E. Schapire. Gambling in a rigged casino: The adversarial multi-armed bandit problem. In *Proceedings of IEEE 36th Annual Foundations of Computer Science*, pp. 322–331, 1995. doi: 10.1109/SFCS.1995.492488.
- P. Auer, N. Cesa-Bianchi, and Fischer P. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47:235–256, 2002. URL <https://doi.org/10.1023/A:1013689704352>.
- Martin J. Bailey. The demand revealing process: To distribute the surplus. *Public Choice*, 91(2):107–126, 1997. ISSN 00485829, 15737101. URL <http://www.jstor.org/stable/30024208>.
- Maria-Florina Balcan, Tuomas Sandholm, and Ellen Vitercik. Sample complexity of automated mechanism design. In *Proceedings of the 30th International Conference on Neural Information Processing Systems, NIPS’16*, pp. 2091–2099, Red Hook, NY, USA, 2016. Curran Associates Inc. ISBN 9781510838819.
- Sébastien Bubeck, Rémi Munos, and Gilles Stoltz. Pure exploration in multi-armed bandits problems. In Ricard Gavaldà, Gábor Lugosi, Thomas Zeugmann, and Sandra Zilles (eds.), *Algorithmic Learning Theory*, pp. 23–37, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg. ISBN 978-3-642-04414-4.
- Ruggiero Cavallo. Optimal decision-making with minimal waste: Strategyproof redistribution of VCG payments. In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS ’06*, pp. 882–889, New York, NY, USA, 2006. Association for Computing Machinery. ISBN 1595933034. doi: 10.1145/1160633.1160790. URL <https://doi.org/10.1145/1160633.1160790>.
- Herman Chernoff. *Sequential Analysis and Optimal Design*. CBMS-NSF Regional Conference Series in Applied Mathematics, Series Number 8. Society for Industrial and Applied Mathematics, 1972.
- Sergei Chichin, Quoc Bao Vo, and Ryszard Kowalczyk. Towards efficient and truthful market mechanisms for double-sided cloud markets. *IEEE Transactions on Services Computing*, 10(1):37–51, 2017. doi: 10.1109/TSC.2016.2594764.
- Vincent Conitzer and Tuomas Sandholm. Complexity of mechanism design. In *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence*, pp. 103–110, August 2002.
- Michael Curry, Ping-Yeh Chiang, Tom Goldstein, and John Dickerson. Certifying strategyproof auction networks. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 4987–4998. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/3465ab6e0c21086020e382f09a482ced-Paper.pdf.
- Paul Duetting, Zhe Feng, Harikrishna Narasimhan, David Parkes, and Sai Srivatsa Ravindranath. Optimal auctions through deep learning. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 1706–1715. PMLR, 09–15 Jun 2019. URL <https://proceedings.mlr.press/v97/duetting19a.html>.
- Lachlan Dufton, Victor Naroditskiy, Maria Polukarov, and Nicholas Jennings. Optimizing payments in dominant-strategy mechanisms for multi-parameter domains. *Proceedings of the AAAI Conference on Artificial Intelligence*, 26(1):1347–1354, Sep. 2021. doi: 10.1609/aaai.v26i1.8262. URL <https://ojs.aaai.org/index.php/AAAI/article/view/8262>.

-
- 594 Eyal Even-Dar, Shie Mannor, and Yishay Mansour. PAC bounds for Multi-armed bandit and Markov
595 decision processes. In Jyrki Kivinen and Robert H. Sloan (eds.), *Computational Learning Theory*,
596 pp. 255–270, Berlin, Heidelberg, 2002. Springer Berlin Heidelberg. ISBN 978-3-540-45435-9.
597
- 598 Eyal Even-Dar, Shie Mannor, and Yishay Mansour. Action elimination and stopping conditions
599 for the multi-armed bandit and reinforcement learning problems. *Journal of Machine Learning*
600 *Research*, 7(39):1079–1105, 2006. URL [http://jmlr.org/papers/v7/evendar06a.](http://jmlr.org/papers/v7/evendar06a.html)
601 [html](http://jmlr.org/papers/v7/evendar06a.html).
- 602 J. R. Green and J. J. Laffont. Characterization of satisfactory mechanism for the revelation of
603 preferences for public goods. *Econometrica*, 45:427–438, 1977.
604
- 605 Jerry R. Green and Jean-Jacques Laffont. *Incentives in Public Decision Making*. North-Holland
606 Publishing Company, Amsterdam, 1979.
607
- 608 Sujit Gujar and Y. Narahari. Redistribution mechanisms for assignment of heterogeneous objects.
609 *Journal of Artificial Intelligence Research*, 41(2):131–154, May 2011. ISSN 1076-9757.
- 610 Mingyu Guo. VCG redistribution with gross substitutes. *Proceedings of the AAAI Conference on*
611 *Artificial Intelligence*, 25(1):675–680, Aug. 2011. doi: 10.1609/aaai.v25i1.7883. URL [https:](https://ojs.aaai.org/index.php/AAAI/article/view/7883)
612 [/ojs.aaai.org/index.php/AAAI/article/view/7883](https://ojs.aaai.org/index.php/AAAI/article/view/7883).
613
- 614 Mingyu Guo. Worst-case optimal redistribution of VCG payments in heterogeneous-item auctions
615 with unit demand. In *Proceedings of the 11th International Conference on Autonomous Agents*
616 *and Multiagent Systems - Volume 2, AAMAS '12*, pp. 745–752, Richland, SC, 2012. International
617 Foundation for Autonomous Agents and Multiagent Systems. ISBN 0981738125.
- 618 Mingyu Guo and Vincent Conitzer. Worst-case optimal redistribution of VCG payments. In
619 *Proceedings of the 8th ACM Conference on Electronic Commerce, EC '07*, pp. 30–39, New
620 York, NY, USA, 2007. Association for Computing Machinery. ISBN 9781595936530. doi:
621 10.1145/1250910.1250915. URL <https://doi.org/10.1145/1250910.1250915>.
622
- 623 Mingyu Guo and Vincent Conitzer. Worst-case optimal redistribution of VCG payments in multi-
624 unit auctions. *Games and Economic Behavior*, 67(1):69–98, 2009. ISSN 0899-8256. doi:
625 <https://doi.org/10.1016/j.geb.2008.06.007>. URL [https://www.sciencedirect.com/](https://www.sciencedirect.com/science/article/pii/S0899825608001267)
626 [science/article/pii/S0899825608001267](https://www.sciencedirect.com/science/article/pii/S0899825608001267). Special Section of Games and Economic
627 Behavior Dedicated to the 8th ACM Conference on Electronic Commerce.
- 628 Avinatan Hassidim, Ron Kupfer, and Yaron Singer. An optimal elimination algorithm for learn-
629 ing a best arm. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.),
630 *Advances in Neural Information Processing Systems*, volume 33, pp. 10788–10798. Curran
631 Associates, Inc., 2020. URL [https://proceedings.neurips.cc/paper_files/](https://proceedings.neurips.cc/paper_files/paper/2020/file/7a43ed4e82d06a1e6b2e88518fb8c2b0-Paper.pdf)
632 [paper/2020/file/7a43ed4e82d06a1e6b2e88518fb8c2b0-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2020/file/7a43ed4e82d06a1e6b2e88518fb8c2b0-Paper.pdf).
- 633 John William Hatfield, Scott Duke Kominers, Alexandru Nichifor, Michael Ostrovsky, and Alexan-
634 der Westkamp. Stability and competitive equilibrium in trading networks. *Journal of Political*
635 *Economy*, 121(5):966–1005, 2013.
636
- 637 Benjamin F. Hobbs, Michael H. Rothkopf, Laurel C. Hyde, and Richard P. O’Neill. Evalua-
638 tion of a truthful revelation auction in the context of energy markets with nonconcave benefits.
639 *Journal of Regulatory Economics*, 18:5–32, 2000. URL [https://doi.org/10.1023/A:](https://doi.org/10.1023/A:1008151712010)
640 [1008151712010](https://doi.org/10.1023/A:1008151712010).
- 641 Dmitry Ivanov, Iskander Safiulin, Igor Filippov, and Ksenia Balabaeva. Optimal-er auctions through
642 attention. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho (eds.), *Ad-*
643 *vances in Neural Information Processing Systems*, 2022. URL [https://openreview.net/](https://openreview.net/forum?id=Xa1T165JEhB)
644 [forum?id=Xa1T165JEhB](https://openreview.net/forum?id=Xa1T165JEhB).
645
- 646 Matthew O. Jackson. Mechanism theory, December 2014. URL [http://dx.doi.org/10.](http://dx.doi.org/10.2139/ssrn.2542983)
647 [2139/ssrn.2542983](http://dx.doi.org/10.2139/ssrn.2542983). Available at SSRN: [http://dx.doi.org/10.2139/ssrn.](http://dx.doi.org/10.2139/ssrn.2542983)
[2542983](http://dx.doi.org/10.2139/ssrn.2542983).

- 648 Hiroshi Kajino, Kohei Miyaguchi, and Takayuki Osogami. Biases in evaluation of molecular opti-
649 mization methods and bias reduction strategies. In Andreas Krause, Emma Brunskill, Kyunghyun
650 Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th In-*
651 *ternational Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning*
652 *Research*, pp. 15567–15585. PMLR, 23–29 Jul 2023. URL [https://proceedings.mlr.](https://proceedings.mlr.press/v202/kajino23a.html)
653 [press/v202/kajino23a.html](https://proceedings.mlr.press/v202/kajino23a.html).
- 654 Dinesh Kumar, Gaurav Baranwal, Zahid Raza, and Deo Prakash Vidyarthi. A truthful combi-
655 natorial double auction-based marketplace mechanism for cloud computing. *Journal of Sys-*
656 *tems and Software*, 140:91–108, 2018. ISSN 0164-1212. doi: [https://doi.org/10.1016/j.jss.](https://doi.org/10.1016/j.jss.2018.03.003)
657 [2018.03.003](https://doi.org/10.1016/j.jss.2018.03.003). URL [https://www.sciencedirect.com/science/article/pii/](https://www.sciencedirect.com/science/article/pii/S0164121218300451)
658 [S0164121218300451](https://www.sciencedirect.com/science/article/pii/S0164121218300451).
- 659 Tor Lattimore and Csaba Szepesvári. *Bandit Algorithms*. Cambridge University Press, September
660 2020.
- 661 Padala Manisha, C. V. Jawahar, and Sujit Gujar. Learning optimal redistribution mechanisms
662 through neural networks. In *Proceedings of the 17th International Conference on Autonomous*
663 *Agents and MultiAgent Systems*, pp. 345–353, July 2018.
- 664 Shie Mannor and John N. Tsitsiklis. The sample complexity of exploration in the multi-armed bandit
665 problem. *Journal of Machine Learning Research*, 5:623–648, December 2004. ISSN 1532-4435.
- 666 Oded Maron and Andrew Moore. Hoeffding races: Accelerating model selection search
667 for classification and function approximation. In J. Cowan, G. Tesauro, and J. Alspecter
668 (eds.), *Advances in Neural Information Processing Systems*, volume 6. Morgan-Kaufmann,
669 1993. URL [https://proceedings.neurips.cc/paper_files/paper/1993/](https://proceedings.neurips.cc/paper_files/paper/1993/file/02a32ad2669e6fe298e607fe7cc0e1a0-Paper.pdf)
670 [file/02a32ad2669e6fe298e607fe7cc0e1a0-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/1993/file/02a32ad2669e6fe298e607fe7cc0e1a0-Paper.pdf).
- 671 Volodymyr Mnih, Csaba Szepesvári, and Jean-Yves Audibert. Empirical Bernstein stopping. In
672 *Proceedings of the 25th International Conference on Machine Learning, ICML ’08*, pp. 672–679,
673 New York, NY, USA, 2008. Association for Computing Machinery. ISBN 9781605582054. doi:
674 [10.1145/1390156.1390241](https://doi.org/10.1145/1390156.1390241). URL <https://doi.org/10.1145/1390156.1390241>.
- 675 Jamie H Morgenstern and Tim Roughgarden. On the pseudo-dimension of nearly opti-
676 mal auctions. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett (eds.),
677 *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc.,
678 2015. URL [https://proceedings.neurips.cc/paper_files/paper/2015/](https://proceedings.neurips.cc/paper_files/paper/2015/file/fbd7939d674997cdb4692d34de8633c4-Paper.pdf)
679 [file/fbd7939d674997cdb4692d34de8633c4-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2015/file/fbd7939d674997cdb4692d34de8633c4-Paper.pdf).
- 680 Hervé Moulin. Almost budget-balanced VCG mechanisms to assign multiple objects. *Journal of*
681 *Economic Theory*, 144(1):96–119, 2009. ISSN 0022-0531. doi: [https://doi.org/10.1016/j.jet.](https://doi.org/10.1016/j.jet.2008.03.003)
682 [2008.03.003](https://doi.org/10.1016/j.jet.2008.03.003). URL [https://www.sciencedirect.com/science/article/pii/](https://www.sciencedirect.com/science/article/pii/S0022053108000628)
683 [S0022053108000628](https://www.sciencedirect.com/science/article/pii/S0022053108000628).
- 684 R. B. Myerson. Optimal auction design. *Mathematics of Operations Research*, 6:58–73, 1981.
- 685 Roger B. Myerson and Mark A. Satterthwaite. Efficient mechanisms for bilateral trading. *Journal*
686 *of Economic Theory*, 29(2):265–281, 1983.
- 687 Noam Nisan. Introduction to mechanism design (for computer scientists). In Noam Nisan, Tim
688 Roughgarden, Eva Tardos, and Vijay V. Vazirani (eds.), *Algorithmic Game Theory*, pp. 209–242.
689 Cambridge University Press, 2007.
- 690 Takayuki Osogami, Segev Wasserkrug, and Elisheva S. Shamash. Learning efficient truthful mecha-
691 nisms for trading networks. In Edith Elkind (ed.), *Proceedings of the Thirty-Second International*
692 *Joint Conference on Artificial Intelligence, IJCAI-23*, pp. 2862–2869. International Joint Con-
693 *ferences on Artificial Intelligence Organization*, 8 2023. doi: 10.24963/ijcai.2023/319. URL
694 <https://doi.org/10.24963/ijcai.2023/319>. Main Track.
- 695 Jad Rahme, Samy Jelassi, and S. Matthew Weinberg. Auction learning as a two-player game. In
696 *International Conference on Learning Representations*, 2021. URL [https://openreview.](https://openreview.net/forum?id=YHdeAO6116T)
697 [net/forum?id=YHdeAO6116T](https://openreview.net/forum?id=YHdeAO6116T).

702 Tuomas Sandholm. Automated mechanism design: A new application area for search algorithms. In
703 Francesca Rossi (ed.), *Principles and Practice of Constraint Programming – CP 2003*, pp. 19–36,
704 Berlin, Heidelberg, 2003. Springer Berlin Heidelberg. ISBN 978-3-540-45193-8.

705

706 Yoav Shoham and Kevin Leyton-Brown. *Multiagent Systems: Algorithmic, Game-Theoretic,*
707 *and Logical Foundations*. Cambridge University Press, 2009. URL [http://www.](http://www.masfoundations.org/mas.pdf)
708 [masfoundations.org/mas.pdf](http://www.masfoundations.org/mas.pdf).

709 Jochen Stöber, Dirk Neumann, and Christof Weinhardt. Market-based pricing in grids: On strategic
710 manipulation and computational cost. *European Journal of Operational Research*, 203(2):464–
711 475, 2010. ISSN 0377-2217. doi: <https://doi.org/10.1016/j.ejor.2009.07.033>. URL [https://](https://www.sciencedirect.com/science/article/pii/S0377221709005323)
712 www.sciencedirect.com/science/article/pii/S0377221709005323.

713 Vasilis Syrgkanis. A sample complexity measure with applications to learning optimal auc-
714 tions. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan,
715 and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Cur-
716 ran Associates, Inc., 2017. URL [https://proceedings.neurips.cc/paper_files/](https://proceedings.neurips.cc/paper_files/paper/2017/file/a7a3d70c6d17a73140918996d03c014f-Paper.pdf)
717 [paper/2017/file/a7a3d70c6d17a73140918996d03c014f-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2017/file/a7a3d70c6d17a73140918996d03c014f-Paper.pdf).

718

719 Andrea Tacchetti, DJ Strouse, Marta Garnelo, Thore Graepel, and Yoram Bachrach. Learning
720 truthful, efficient, and welfare maximizing auction rules. In *ICLR 2022 Workshop on Gami-*
721 *fication and Multiagent Solutions*, 2022. URL [https://openreview.net/forum?id=](https://openreview.net/forum?id=Sql8oqJTe9)
722 [Sql8oqJTe9](https://openreview.net/forum?id=Sql8oqJTe9).

723 Hado van Hasselt. Double Q-learning. In J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel,
724 and A. Culotta (eds.), *Advances in Neural Information Processing Systems*, volume 23. Cur-
725 ran Associates, Inc., 2010. URL [https://proceedings.neurips.cc/paper_files/](https://proceedings.neurips.cc/paper_files/paper/2010/file/091d584fced301b442654dd8c23b3fc9-Paper.pdf)
726 [paper/2010/file/091d584fced301b442654dd8c23b3fc9-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2010/file/091d584fced301b442654dd8c23b3fc9-Paper.pdf).

727

728 Hado van Hasselt. Estimating the maximum expected value: An analysis of (nested) cross validation
729 and the maximum sample average, 2013. URL <https://arxiv.org/abs/1302.7175>.
730 arXiv:1302.7175.

731 Segev Wasserkrug and Takayuki Osogami. Who benefits from a multi-cloud market? A trading
732 networks based analysis, 2023. arXiv:2310.12666.

733 Tobias Widmer and Joerg Leukel. Efficiency of electronic service allocation with privately known
734 quality. *European Journal of Operational Research*, 255(3):856–868, 2016. ISSN 0377-2217.
735 doi: <https://doi.org/10.1016/j.ejor.2016.05.055>. URL [https://www.sciencedirect.](https://www.sciencedirect.com/science/article/pii/S0377221716303927)
736 [com/science/article/pii/S0377221716303927](https://www.sciencedirect.com/science/article/pii/S0377221716303927).

737

738 Tao Zhang and Quanyan Zhu. Optimal two-sided market mechanism design for large-scale data
739 sharing and trading in massive IoT networks, 2020. arXiv:1912.06229.

740 Xiaoyan Zou. Double-sided auction mechanism design in electricity based on maximizing social
741 welfare. *Energy Policy*, 37(11):4231–4239, 2009. ISSN 0301-4215. doi: [https://doi.org/10.1016/](https://doi.org/10.1016/j.enpol.2009.05.019)
742 [j.enpol.2009.05.019](https://doi.org/10.1016/j.enpol.2009.05.019). URL [https://www.sciencedirect.com/science/article/](https://www.sciencedirect.com/science/article/pii/S0301421509003437)
743 [pii/S0301421509003437](https://www.sciencedirect.com/science/article/pii/S0301421509003437).

744

745

746

747

748

749

750

751

752

753

754

755

756 A DETAILS

757
758 In this section, we provide full details of derivations and other details skipped in the body of the
759 paper.
760

761 A.1 DETAILS OF SECTION 4

762 Equivalence in (8) follows from

$$764 \mathbb{E}[v(\phi^*(t); t_n) - \tau_n(t) \mid t_n] \geq \theta(t_n)$$

$$765 \iff \mathbb{E} \left[\sum_{m \in \mathcal{N}} v(\phi^*(t); t_m) \mid t_n \right] - \mathbb{E}[h_n(t_{-n}) \mid t_n] \geq \theta(t_n) \quad (28)$$

$$767 \iff \mathbb{E}[w^*(t) \mid t_n] - \mathbb{E}[h_n(t_{-n}) \mid t_n] \geq \theta(t_n). \quad (29)$$

770 Equivalence in (9) follows from

$$771 \sum_{n \in \mathcal{N}} \mathbb{E}[\tau_n(t)] \geq \rho \iff \sum_{n \in \mathcal{N}} \mathbb{E}[h_n(t_{-n})] - \sum_{n \in \mathcal{N}} \sum_{m \in \mathcal{N}_{-n}} \mathbb{E}[v(\phi^*(t); t_m)] \geq \rho \quad (30)$$

$$772 \iff \sum_{n \in \mathcal{N}} \mathbb{E}[h_n(t_{-n})] - (N-1) \mathbb{E}[w^*(t)] \geq \rho, \quad (31)$$

773 where the last equivalence follows from the definition of $w^*(t)$ in (7).
774

775 **Proposition 4.** *Let the decision rule ϕ^* be the one that satisfies (3) and the payment rule $\tau = (\tau_n)_n$
776 be in the form of (6) where $h = (h_n)_n = (h_n^*)_n = h^*$ is given by the solution to the LP (10)-(12).
777 Then the VCG mechanism (ϕ^*, τ^*) satisfies DSIC, DE, θ -IR, and ρ -WBB.*
778

779 *Proof.* With the equivalences (8)-(9), the constraints (11)-(12) in the LP guarantee that θ -IR and
780 ρ -WBB are satisfied by feasible solutions. Since we consider the class of VCG mechanisms, DE is
781 trivially satisfied by the definition of ϕ^* , and DSIC is satisfied when the payment rule is in the form
782 of (6). Hence, all of DSIC, DE, θ -IR, and ρ -WBB are satisfied by (ϕ^*, θ^*) . \square
783

784 Algorithm 1 summarizes the protocol under the VCG mechanism discussed Section 4. In Step 3, the
785 optimal strategy of each player is to truthfully declare its type $\hat{t}_n = t_n$. In Step 5, the LP may not be
786 feasible, in which case the protocol may fail, or we may use another payment rule to proceed.
787

788 **Algorithm 1** Protocol under the VCG mechanism

- 791 1: Sample the type profile t from the common prior \mathbb{P}
 - 792 2: Each player $n \in \mathcal{N}$ gets to know its own type t_n
 - 793 3: Each player $n \in \mathcal{N}$ declare their type \hat{t}_n
 - 794 4: Determine the social decision: $\phi^*(\hat{t}) \leftarrow \operatorname{argmax}_{d \in \mathcal{D}} \sum_{n \in \mathcal{N}} v(d; \hat{t}_n)$
 - 795 5: $h^* \leftarrow$ Find the optimal solution to the LP (10)-(12)
 - 796 6: Determine the payment from each player $n \in \mathcal{N}$ to the mediator:
797 $\tau_n(\hat{t}) \leftarrow h_n^*(\hat{t}_{-n}) - \sum_{m \in \mathcal{N}_{-n}} v(\phi^*(\hat{t}); \hat{t}_m)$
 - 798 7: Each player $n \in \mathcal{N}$ gets utility $v(\phi^*(\hat{t}); t_n) - \tau_n(\hat{t})$
-

801 A.2 DETAILS OF SECTION 5

802 A.2.1 SOLUTIONS THAT GUARANTEE WBB

803 Alternatively, one may set

$$804 \theta(t_n) = \left[\mathbb{E}[w^*(t) \mid t_n] - \frac{N-1}{N} \mathbb{E}[w^*(t)] \right]^- \quad \forall t_n \in \mathcal{T}_n, \forall n \in \mathcal{N} \quad (32)$$

$$805 \rho = 0 \quad (33)$$

Algorithm 2 PAC-BAI(ε, δ, T)

- 1: Run an arbitrary (ε, δ) -PAC BAI with sample complexity T
 - 2: Let \hat{I} be the ε -optimal arm identified by the algorithm
 - 3: Let $M_{\hat{I}}$ be the number of times the algorithm has pulled arm \hat{I} ; if no such information is available, set $M_{\hat{I}} = 0$
 - 4: Let $\hat{\mu}_{\hat{I}}$ be the sample average of the $M_{\hat{I}}$ rewards obtained from arm \hat{I} ; if $M_{\hat{I}} = 0$, set $\hat{\mu}_{\hat{I}} = 0$
 - 5: **return** $\hat{I}, M_{\hat{I}}, \hat{\mu}_{\hat{I}}$
-

Algorithm 3 Best Mean Estimator

- 1: $\hat{I}, M_{\hat{I}}, \hat{\mu}_{\hat{I}} \leftarrow \text{PAC-BAI}(\varepsilon, \delta, T)$
 - 2: **if** $M_{\hat{I}} < m^*$, where m^* is defined in Lemma 4 **then**
 - 3: Pull arm \hat{I} for $m^* - M_{\hat{I}}$ times
 - 4: Update the sample average $\hat{\mu}_{\hat{I}}$ of arm \hat{I}
 - 5: **end if**
 - 6: **return** $\hat{\mu}_{\hat{I}}$
-

to guarantee the feasibility, since

$$\begin{aligned} & \sum_{n \in \mathcal{N}} \min_{t_n \in \mathcal{T}_n} \left\{ \mathbb{E}[w | t_n] - \left[\mathbb{E}[w^*(t) | t_n] - \frac{N-1}{N} \mathbb{E}[w^*(t)] \right]^- \right\} - (N-1) \mathbb{E}[w^*(t)] \\ &= \sum_{n \in \mathcal{N}} \left(\min_{t_n \in \mathcal{T}_n} \left\{ \mathbb{E}[w | t_n] - \frac{N-1}{N} \mathbb{E}[w^*(t)] - \left[\mathbb{E}[w^*(t) | t_n] - \frac{N-1}{N} \mathbb{E}[w^*(t)] \right]^- \right\} \right) \quad (34) \\ &\geq 0. \quad (35) \end{aligned}$$

In this case, player n may incur negative utility when it has type t_n with $\theta(t_n) < 0$, although the loss is guaranteed to be bounded by $|\theta(t_n)|$.

A.2.2 ON DEPENDENT TYPES

The condition (13) may be satisfied even when types are dependent, and the optimality of our analytic solutions in Lemma 3 is guaranteed as long as (13) is satisfied. When types are dependent, however, there are cases where feasible solutions exist even when (13) is violated (Proposition 1).

In the proof of Proposition 1, we construct such a case with an extreme example of completely dependent types. However, (13) is often satisfied even in such extreme cases of completely dependent types. For example, as long as

$$x_1 \leq x_2 \leq \frac{2-p}{1-p} x_1, \quad (36)$$

condition (13) is satisfied in the example in the proof of Proposition 1, since then (x_1, x_2) satisfies (63), which corresponds to (13) in this example.

A.3 DETAILS OF SECTION 6

A.3.1 UPPER BOUND

For effective use of sample, we consider Algorithm 2, which wraps an arbitrary (ε, δ) -PAC BAI. Specifically, Algorithm 2 not only returns the arm \hat{I} that is identified as best by the (ε, δ) -PAC BAI but also returns the number $M_{\hat{I}}$ of i.i.d. samples that the (ε, δ) -PAC BAI has taken from \hat{I} as well as the corresponding sample average $\hat{\mu}_{\hat{I}}$. However, if no such information is available, it is perfectly fine, and Algorithm 2 simply returns $M_{\hat{I}} = 0$ and $\hat{\mu}_{\hat{I}} = 0$ in such a case.

Notice that the naive approach of sampling each arm $\Theta((1/\varepsilon^2) \log(1/\delta))$ times, which also trivially falls within the upper bound in Theorem 1, would only guarantee that the best mean is estimated

864 with the error bound of ε with probability at least $(1 - \delta)^K$. Conversely, it would require sampling
865 each arm $\Omega((1/\varepsilon^2) \log(K/\delta))$ times to obtain the same error bound with probability $1 - \delta$.
866

867 A.3.2 LOWER BOUND 868

869 The lower bound established in Lemma 4 reduces BME to the ε -Biased Coin Problem defined as
870 follows:

871 **Definition 3** (ε -Biased Coin Problem). *For $0 < \varepsilon < 1$, consider a Bernoulli random variable B
872 whose mean α is known to be either $\alpha^+ := (1 + \varepsilon)/2$ or $\alpha^- := (1 - \varepsilon)/2$. The ε -Biased Coin
873 Problem asks to correctly identify whether $\alpha = \alpha^+$ or $\alpha = \alpha^-$.*
874

875 A lower bound on the sample complexity for solving the ε -Biased Coin Problem is known as the
876 following lemma, for which we provide a proof in Appendix B for completeness:

877 **Lemma 7** (Chernoff (1972); Even-Dar et al. (2002)). *For $0 < \delta < 1/2$, any algorithm that solves
878 the ε -Biased Coin Problem correctly with probability at least $1 - \delta$ has expected sample complexity
879 at least $\Omega((1/\varepsilon^2) \log(1/\delta))$.*
880

881 In Lemma 5, we have derived the lower bound for BME using the technique used for a lower bound
882 for BAI in Even-Dar et al. (2002); however, as we have discussed at the end of Section 6, while
883 our lower bound for BME is tight, the lower bound for BAI in Even-Dar et al. (2002) is not. The
884 difference in the derived lower bound stems from the following behavior of BME and BAI when all
885 of the arms have mean reward of α^- and hence are indistinguishable. The algorithm constructed in
886 Even-Dar et al. (2002) determines that B has mean α^+ when either arm i^+ or arm i^- is identified
887 as the best arm. When the arms are indistinguishable, a PAC BAI would correctly identify each
888 of the K arms, including i^+ or i^- , as the best arm uniformly at random, which induces an error
889 with probability $1/K$. On the other hand, the mean reward estimated by a PAC BME would be
890 approximately correct with high probability, even when the arms are indistinguishable.

891 A.4 DETAILS OF SECTION 7 892

893 In this section, we conduct several numerical experiments to validate the effectiveness of the pro-
894 posed approach and to understand its limitations. Specifically, we design our experiments to inves-
895 tigate the following three questions:

- 896 1. BME studied in Section 6 has asymptotically optimal sample complexity, but how well can
897 we estimate the best mean when there are only a moderate number of arms?
- 898 2. How much can BME reduce the number of times $w^*(t)$ given by (7) is evaluated?
- 899 3. How well θ -IR and ρ -SBB are satisfied when (20) is estimated by BME rather than calcu-
900 lated exactly?
901

902 A.4.1 BEST MEAN ESTIMATION WITH MODERATE NUMBER OF ARMS 903

904 For BAI, existing algorithms that have asymptotically optimal sample complexity often perform
905 poorly with moderate number of arms (Hassidim et al., 2020). As a result, Approximate Best Arm
906 (Hassidim et al., 2020), which has optimal sample complexity, runs Naive Elimination, which has
907 asymptotically suboptimal sample complexity, when the number of arms is below 10^5 or after suf-
908 ficient number of suboptimal arms is eliminated. The BME algorithm studied in Section 6 has
909 asymptotically optimal sample complexity but relies on BAI, and thus its sample complexity for
910 moderate number of arms is not well characterized by Theorem 1. Similar to BAI, BME relies on
911 an algorithm that performs well when there are only moderate number of arms.

912 In this section, we compare the performance of Successive Elimination algorithms for BAI and
913 BME. Successive Elimination has suboptimal sample complexity, similar to Naive Elimination, but
914 often outperforms Naive Elimination for moderate number of arms (Even-Dar et al., 2006). Specif-
915 ically, we compare the performance of (ε, δ) -PAC Successive Elimination for BME (SE-BME; Al-
916 gorithm 4) against the corresponding (ε, δ) -PAC Successive Elimination for BAI (SE-BAI; Algo-
917 rithm 5). For completeness, in Appendix B, we provide standard proofs on the correctness of these
algorithms, as stated in Proposition 5 and Proposition 6:

Algorithm 4 Successive Elimination for Best Mean Estimation $((\varepsilon, \delta)$ -PAC SE-BME)

Require: ε, δ

- 1: Let $\mathcal{R} \leftarrow \{1, \dots, K\}$ be the set of remaining arms
 - 2: Let $t \leftarrow 0; \alpha \leftarrow 1$
 - 3: **while** $\alpha > \varepsilon$ **do**
 - 4: Pull each arm $k \in \mathcal{R}$ once and update the sample average $\hat{\mu}_k$
 - 5: $t \leftarrow t + 1$
 - 6: $\alpha \leftarrow \sqrt{\frac{1}{2t} \log\left(\frac{\pi^2 K t^2}{3\delta}\right)}$
 - 7: **for all** $k \in \mathcal{R}$ **do**
 - 8: Remove k from \mathcal{R} if $\max_{\ell \in \mathcal{R}} \hat{\mu}_\ell - \hat{\mu}_k \geq 2\alpha$
 - 9: **end for**
 - 10: **end while**
 - 11: **return** $\max_{k \in \mathcal{R}} \hat{\mu}_k$
-

Algorithm 5 Successive Elimination for Best Arm Identification $((\varepsilon, \delta)$ -PAC SE-BAI)

Require: ε, δ

- 1: Let $\mathcal{R} \leftarrow \{1, \dots, K\}$ be the set of remaining arms
 - 2: Let $t \leftarrow 0; \alpha \leftarrow 1$
 - 3: **while** $|\mathcal{R}| > 1$ and $\alpha > \frac{\varepsilon}{2}$ **do**
 - 4: Pull each arm $k \in \mathcal{R}$ once and update the sample average $\hat{\mu}_k$
 - 5: $t \leftarrow t + 1$
 - 6: $\alpha \leftarrow \sqrt{\frac{1}{2t} \log\left(\frac{\pi^2 K t^2}{6\delta}\right)}$
 - 7: **for all** $k \in \mathcal{R}$ **do**
 - 8: Remove k from \mathcal{R} if $\max_{\ell \in \mathcal{R}} \hat{\mu}_\ell - \hat{\mu}_k \geq 2\alpha$
 - 9: **end for**
 - 10: **end while**
 - 11: **return** $\operatorname{argmax}_{k \in \mathcal{R}} \hat{\mu}_k$
-

Proposition 5. *Algorithm 4 is an (ε, δ) -PAC BME.*

Proposition 6. *Algorithm 5 is an (ε, δ) -PAC BAI.*

There are three differences between SE-BME and SE-BAI. First, SE-BME exits the while-loop when $\alpha < \varepsilon$ (while SE-BAI needs to wait until $\alpha < \varepsilon/2$), since it can return an overestimated mean of a suboptimal arm as long as the estimated value is within ε from the best mean. Second, SE-BAI exits the while-loop when only one arm remains, since it does not need to precisely estimate the mean of the identified arm. Finally, SE-BAI uses smaller α_t , since it only need to consider one-sided estimation error (underestimation for best arms and overestimation for other arms).

Figure 3 shows the total sample size required by SE-BME and by SE-BAI for varying values of ε and δ , and for varying number K of arms. Here, the arms have Bernoulli rewards, and their means are selected in a way they are equally separated (i.e., $\mu_k = (k - 0.5)/K$ for $k = 1, \dots, K$). For each datapoint, the experiments are repeated 10 times. The standard deviation of the total sample size is plotted but too small to be visible in the figure.

Overall, it can be observed that SE-BME generally requires smaller sample size than SE-BAI, except when there are only a few arms (and the means have large gaps in the setting under consideration). The efficiency of SE-BAI for a small number of arms makes intuitive sense, because the best arm can be identified without estimating the means with high accuracy.

A.4.2 EFFECTIVENESS OF BEST MEAN ESTIMATION IN MECHANISM DESIGN

In this section, we quantitatively validate the effectiveness of BME in reducing the number of times $w^*(t)$ is computed when we evaluate (20). We use (ε, δ) -PAC SE-BME (Algorithm 4) as BME.

To this end, we consider the following mechanism design, motivated by the double-sided auctions for electricity (Zou, 2009; Hobbs et al., 2000), where we have $N = |\mathcal{N}|$ players, each player $n \in \mathcal{N}$

972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025

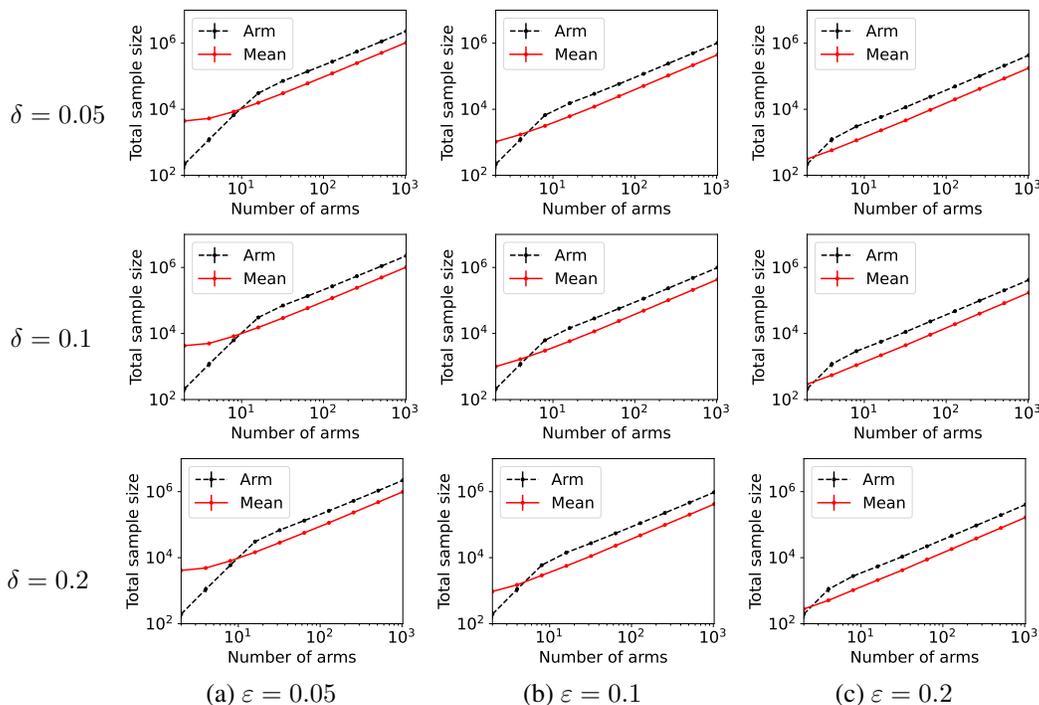


Figure 3: The total sample size required by (ε, δ) -PAC BME (Mean; Algorithm 4) and by (ε, δ) -PAC BAI (Arm; Algorithm 5) when arms have Bernoulli rewards with equally separated means for varying values of ε and δ .

has K possible types (i.e., $|\mathcal{T}_n| = K$), with varying values of N and K . The K possible types of each player are selected uniformly at random from integers, $[-K, K]$, without replacement. We then assume, as the common prior \mathbb{P} , that the type of each player is distributed uniformly among the K possible types and independent of the types of other players. Each player is a buyer or a seller of a single item, depending on its type. When a player n has a positive type t_n , the player is a buyer who wants to buy a unit, whose valuation to the player is t_n (i.e., $v(d; t_n) = t_n$ if player n buys a unit of the item with social decision d). When the player has a negative type t_n , the player is a seller who wants to sell a unit, which incurs cost $|t_n|$ (i.e., $v(d; t_n) = t_n$ if player n sells a unit of the item with social decision d). The player does not participate in the market, when its type is zero. For a given profile of types t , a social decision is given by a bipartite matching between buyers and sellers. In this setting, we can immediately obtain the efficient social decision (3) by greedily matching buyers of high valuations to sellers of low costs as long as the value of the buyers are higher than the costs of the sellers.

Figure 4 shows representative sample paths of the estimated values of $\mathbb{E}[w^*(t) | t_n]$ for each $t_n \in \mathcal{T}_n$ when $(0.25, 0.1)$ -PAC SE-BME¹ is used to evaluate $\min_{t_n \in \mathcal{T}_n} \mathbb{E}[w^*(t) | t_n]$ (i.e., (20) with $\theta(t_n) = 0$). Each panel in Figure 4 shows $|\mathcal{T}_n|$ curves, where the red curve corresponds to the one with minimum $\mathbb{E}[w^*(t) | t_n]$.

Observe that the types t_n (arms) that have close to the minimum mean, $\min_{t_n \in \mathcal{T}_n} \mathbb{E}[w^*(t) | t_n]$, survive until SE-BME terminates, and their means are evaluated with sufficient accuracy. On the other hand, the types that have large means are eliminated after a relatively small number of samples without being estimated precisely, which contributes to reducing the number of evaluating the value of the efficient social decision $w^*(t)$.

In Figure 5, we study how much SE-BME can reduce the number of evaluations of $w^*(t)$ needed to estimate $\min_{t_n \in \mathcal{T}_n} \mathbb{E}[w^*(t) | t]$ for all $n \in \mathcal{N}$ with the accuracy that is shown in Figure 4 (i.e., using the same values of $\varepsilon = 0.25$ and $\delta = 0.1$). Recall that $K = |\mathcal{T}_n|, \forall n \in \mathcal{N}$ in the setting under consideration. Hence, the exact computation of $\min_{t_n \in \mathcal{T}_n} \mathbb{E}[w^*(t) | t_n]$ for a single $n \in \mathcal{N}$

¹Minimization in Algorithm 4 is translated into maximization.

1026
 1027
 1028
 1029
 1030
 1031
 1032
 1033
 1034
 1035
 1036
 1037
 1038
 1039
 1040
 1041
 1042
 1043
 1044
 1045
 1046
 1047
 1048
 1049
 1050
 1051
 1052
 1053
 1054
 1055
 1056
 1057
 1058
 1059
 1060
 1061
 1062
 1063
 1064
 1065
 1066
 1067
 1068
 1069
 1070
 1071
 1072
 1073
 1074
 1075
 1076
 1077
 1078
 1079

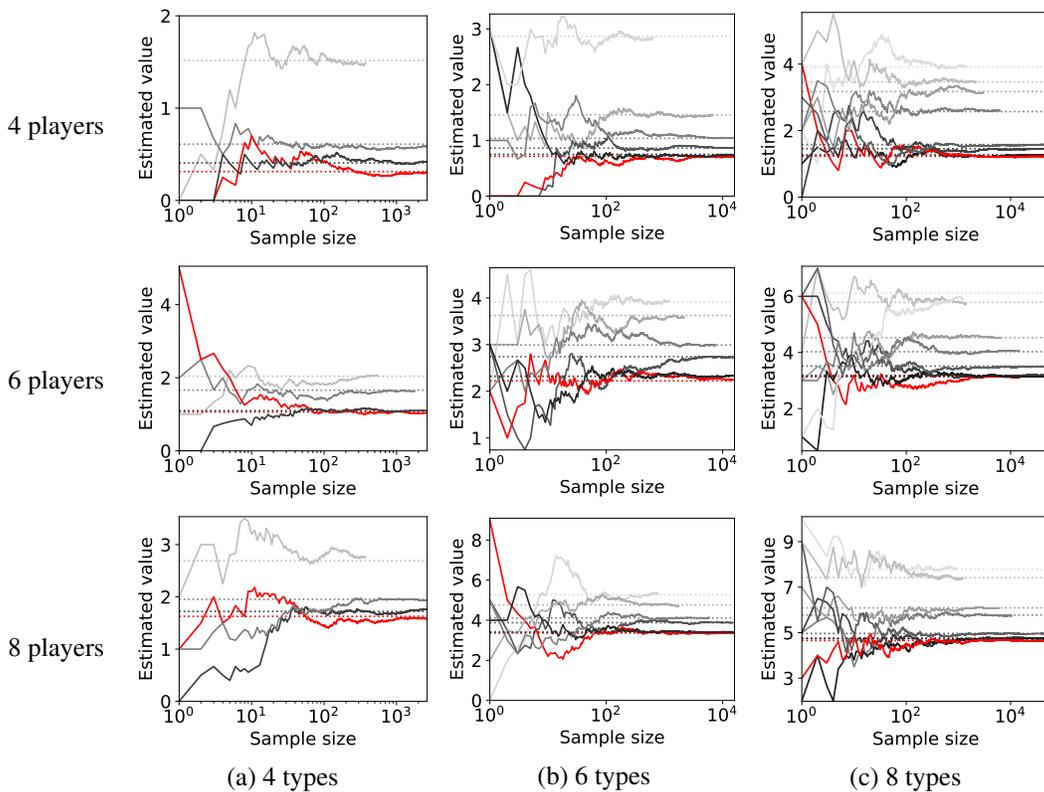


Figure 4: Representative sample paths that show the estimated values of $\mathbb{E}[w^*(t) | t_n]$ for $t_n \in \mathcal{T}_n$ against sample size used by (0.25, 0.1)-PAC SE-BME.

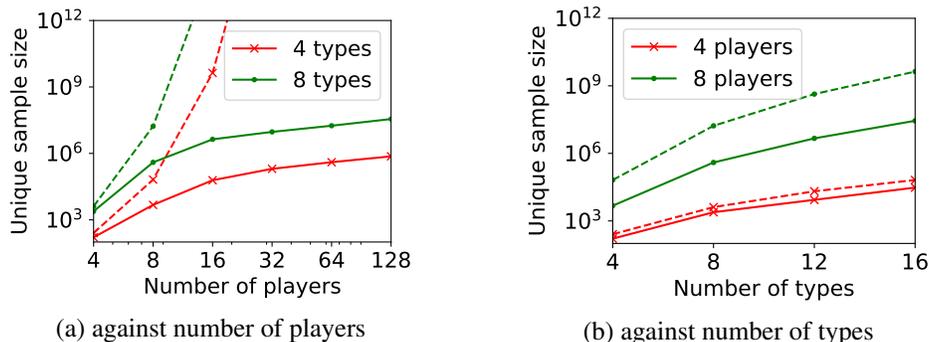
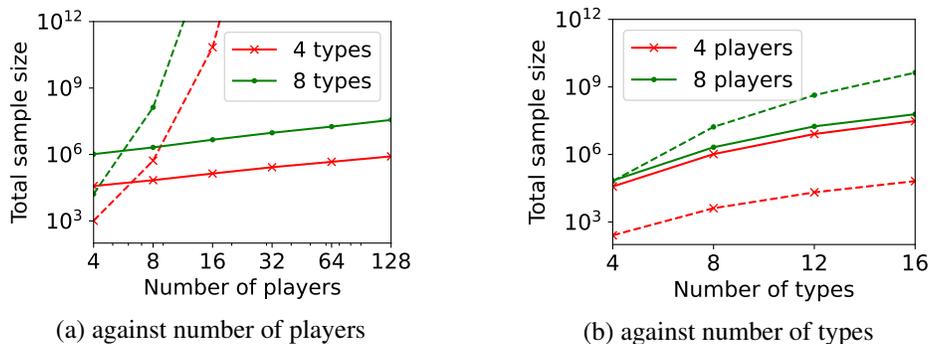


Figure 5: The unique sample size (the number of unique t which $w^*(t)$ is evaluated with) required by the exact computation of $\min_{t_n \in \mathcal{T}_n} \mathbb{E}[w^*(t) | t_n], \forall n \in \mathcal{N}$ (dashed curves) and by (0.25, 0.1)-PAC SE BME (solid curves).

would require evaluating $w^*(t)$ for K^N different values of t , but K^N evaluations of $w^*(t)$ are also sufficient to exactly compute $\min_{t_n \in \mathcal{T}_n} \mathbb{E}[w^*(t) | t_n]$ for all $n \in \mathcal{N}$, because we can cache the value of $w^*(t)$ and reuse it when it is needed. Since the computational complexity associated with evaluating $w^*(t)$ with (7) is the bottleneck, the unique sample size is what we should be interested in. Similar to exact computation, SE-BME also benefits from caching and reusing the values of $w^*(t)$. Figure 5 compares the unique sample size required by the exact computation and (0.25, 0.1)-PAC SE-BME.

Figure 5(a) implies that SE-BME (shown with solid curves) evaluates the total value of the efficient social decision, $w^*(t)$, by orders of magnitude smaller number of times than what is required by

1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091



1092 Figure 6: The total sample size (the number of t which $w^*(t)$ is evaluated with) required by the exact
1093 computation of $\min_{t_n \in \mathcal{T}_n} \mathbb{E}[w^*(t) | t_n], \forall n \in \mathcal{N}$ (dashed curves) and by (0.25, 0.1)-PAC SE-BME
1094 (solid curves).

1095
1096
1097
1098
1099
1100
1101
1102

exact computation (shown with dashed curves). While the number of evaluations of $w^*(t)$ grows exponentially with the number of players (specifically, K^N) when exact computation is used, it grows only polynomially (in fact, slightly slower than linearly) when SE-BME is used. This relative insensitivity of the sample complexity of SE-BME to the number of players makes intuitive sense, because the number of players only affects the distribution of the reward and keeps the number of arms unchanged.

1103
1104
1105
1106
1107

Figure 5(b) shows the number of evaluations of $w^*(t)$ against the number of types $K = |\mathcal{T}_n|$ for any $n \in \mathcal{N}$. The advantage of SE-BME over exact computation is relatively minor when we increase the number of types instead of the number of players, since increasing the number of types directly increases the number of arms. In all cases, however, we can observe that SE-BME can significantly reduce the unique sample size.

1108
1109
1110
1111
1112

Figure 6 shows the total sample size, rather than the unique sample size, required by exact computation (dashed curves) and SE-BME (solid curves). The total sample size with exact computation is $N K^N$. While $w^*(t)$ is evaluated N times for each t with exact computation, SE-BME may waste evaluating the same $w^*(t)$ more often particularly when there are only a small number of players. This reduces benefits of SE-BME for total sample size, as compared to the unique sample size.

1113
1114
1115

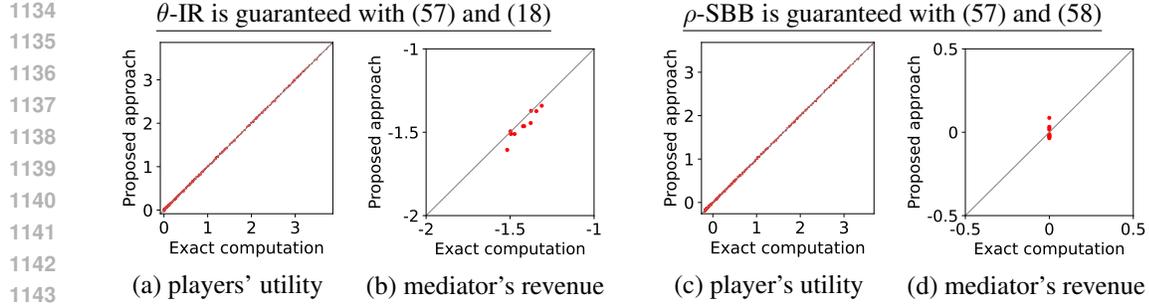
1115 A.4.3 INDIVIDUAL RATIONALITY AND BUDGET BALANCE WITH BEST MEAN ESTIMATION

1116
1117
1118
1119
1120
1121
1122

We next address the question of how well θ -IR and ρ -SBB are guaranteed when we estimate $\min_{t_n \in \mathcal{T}_n} \{\mathbb{E}[w^*(t) | t_n] - \theta(t_n)\}$ with BME rather than computing it exactly. We continue to use the setting of mechanism design introduced in Section A.4.2. Recall that θ -IR is guaranteed when η_n is given by (57) and (18), and ρ -SBB is guaranteed when η_n is given by (57) and (58). However, these are guaranteed only when the expected values, $\mathbb{E}[w^*(t)]$ and $\mathbb{E}[w^*(t) | t_n]$, are exactly computed. In this section, we quantitatively evaluate how well θ -IR and ρ -SBB are satisfied when those expected values are estimated from samples. Throughout this section, we study the case with $\rho = 0$ and $\theta \equiv 0$ (i.e., $\theta(t_n) = 0, \forall t_n \in \mathcal{T}_n, \forall n \in \mathcal{N}$).

1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133

In Figure 7, we first evaluate the best mean, $\min_{t_n \in \mathcal{T}_n} \mathbb{E}[w^*(t) | t]$, either with exact computation or with BME, then compute η_n with (57) and (18) for Columns (a)-(b) and with (57) and (58) for Columns (c)-(d), and finally evaluate the expected utility of each player (the left-hand side of (11)) for Columns (a) and (c) and the expected revenue of the mediator (the left-hand side of (12)) for Columns (b) and (d) by setting $h_n(t_{-n}) = \eta_n, \forall t_{-n} \in \mathcal{T}_{-n}, \forall n \in \mathcal{N}$. Even if the best mean is estimated with BME, we evaluate the expectations on the left-hand side of (11) and (12) with exact computation, because these are the expected utility and the expected revenue that the players and the mediator will experience. Here, we repeat the experiment with 10 different random seeds, so that there are 10 data-points in Columns (b) and (d), and each of Columns (a) and (c) has $10 \times 8 \times 8 = 640$ data-points, where each data-point corresponds to a player of a particular type with a particular random seed.



1145 Figure 7: The red dots show the expected utility of the players in Columns (a) and (c) and the expected revenue of the mediator in Columns (b) and (d), where analytical solutions are evaluated exactly (horizontal axes) or estimated with $(0.25, 0.1)$ -PAC SE-BME (vertical axes) for environments with $|\mathcal{N}| = 8$ players, each having $|\mathcal{T}_n| = 8$ possible types. The analytical solution guarantees θ -IR with $\theta \equiv 0$ in Columns (a) and (b) and ρ -SBB with $\rho = 0$ in Columns (c) and (d). Results are plotted for 10 random seeds. Diagonal lines are also plotted to help understand where the horizontal and vertical axes are equal.

1151

1152

1153

1154

1155 Overall, Columns (a) and (c) of Figure 7 show that the expected utility experienced by the players is relatively insensitive to whether the best mean is evaluated with exact computation (horizontal axes) or with BME (vertical axes). Taking a closer look, we can observe that, in this particular setting, 0 -IR is violated for some of the players in Column (c) even if the best mean is computed exactly, while it is guaranteed for any player of any type in Column (a) if the best mean is computed exactly.

1156

1157

1158

1159

1160 On the other hand, as is shown in Columns (b) and (d), the mediator experiences non-negligible difference in its expected revenue depending on whether the best mean is evaluated with exact computation or with BME. It is to be expected that the mediator experiences relatively larger variance in its expected revenue, because (12) involves the summation $\sum_{n \in \mathcal{N}} \eta_n$, while (11) only involves η_n for a single $n \in \mathcal{N}$. In particular, it is evident in Column (d) that 0 -WBB (let alone 0 -SBB) is violated when the best mean is estimated with BME (vertical axes), while it is always guaranteed with exact computation (horizontal axes). In Column (b), 0 -WBB is violated regardless of whether the best mean is computed exactly or with BME, since satisfying 0 -WBB and 0 -IR for all players (together with DE and DSIC) is impossible in this particular setting.

1161

1162

1163

1164

1165

1166

1167

1168 A simple remedy to this violation of ρ -WBB is to replace the ρ with a $\rho' > \rho$ when we compute the η_n from the best means, $\min_{t_n \in \mathcal{T}_n} \mathbb{E}[w^*(t) | t_n]$ for $n \in \mathcal{N}$, estimated with BME. By considering the (ε, δ) -PAC guarantee for the error in the estimation, we can guarantee that ρ -WBB is satisfied with high probability by setting an appropriate value of ρ' . Analogously, we may replace the $\theta(t_n)$ with a $\theta'(t_n) > \theta(t_n)$ to provide a guaranteed that θ -IR is satisfied with high probability when the best mean is estimated with BME. See Theorem 2.

1169

1170

1171

1172

1173

1174

1175 As an example, we set $\rho' = 0.1$ in Figure 8. The consequence of replacing $\rho = 0$ with $\rho' = 0.1$ is as expected. In Columns (b) and (d), the expected revenue of the mediator when the best mean is estimated with BME (Proposed approach) is shifted to the above by $\rho' - \rho = 0.1$. Although it may be unclear from Columns (a) and (c), the corresponding expected utility of each player is shifted to the left by $(\rho' - \rho)/|\mathcal{N}| = 0.0125$. In practice, we may choose ρ' and θ' by taking into account these shifts as well as the condition on the feasibility of LP (Corollary 1).

1176

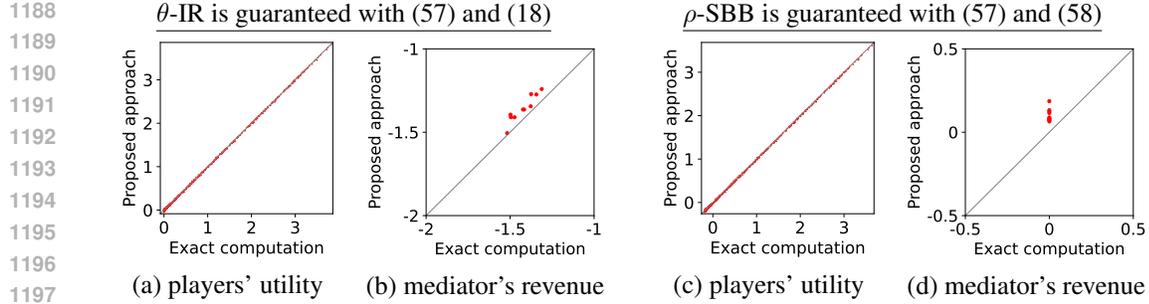
1177

1178

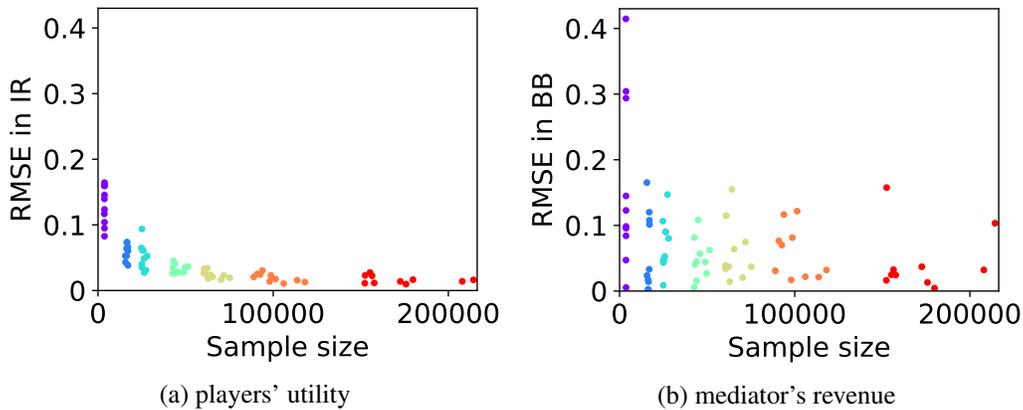
1179

1180

1181 In Figure 9, we show the Root Mean Squared Error (RMSE) in the expected utility of each player (a) and the expected revenue of the mediator (b) that are estimated with BME for the case with $|\mathcal{N}| = 8$ players, each with $|\mathcal{T}_n| = 4$ types. Here, we fix $\delta = 0.1$ and vary ε from 1.0 to 0.15 in the BME. For each pair of (δ, ε) , the experiments are repeated 10 times with different random seeds. The total sample size increases as the value of ε decreases. Hence, the purple dots correspond to $\varepsilon = 1.0$, and the red dots are $\varepsilon = 0.15$. Overall, we can observe that RMSE can be reduced by using small ε at the expense of increased sample size, that relatively large values such as $\varepsilon = 0.5$ gives reasonably small RMSE, and that larger values of ε have diminishing effects on RMSE.



1198 Figure 8: The red dots show the expected utility of the players in Columns (a) and (c) and the expected revenue of the mediator in Columns (b) and (d), where analytical solutions are evaluated exactly (horizontal axes) or estimated with $(0.25, 0.1)$ -PAC SE-BME (vertical axes) for environments with $|\mathcal{N}| = 8$ players, each having $|\mathcal{T}_n| = 8$ possible types. The analytical solution guarantees θ -IR with $\theta \equiv 0$ in Columns (a) and (b) and ρ' -SBB with $\rho' = 0.1$ in Columns (c) and (d). Results are plotted for 10 random seeds. Diagonal lines are also plotted to help understand where the horizontal and vertical axes are equal.



1223 Figure 9: Root mean squared error in (a) the expected utility of each player and (b) the expected revenue of the mediator against the total sample size, when there are $|\mathcal{N}| = 8$ players, each with $|\mathcal{T}_n| = 4$ possible types. Here, we set $\delta = 0.1$ and vary ε from 1.0 (purple), 0.5, 0.4, 0.3, 0.25, 0.2, to 0.15 (red).

1232 A.4.4 COMPUTATIONAL REQUIREMENTS

1233

1234

1235 We have run all of the experiments on a single core with at most 66 GB memory without GPUs in a cloud environment. The associated source code is submitted as a supplementary material and will be open-sourced upon acceptance. Table 1 summarizes the CPU time and maximum memory require to generate each figure. For example, CPU time for Figure 3(a) is the time to generate three panels in Column (a) of Figure 3. Note that the CPU time and maximum memory reported in Table 1 are not optimized and include time and memory for storing intermediate results and other processing for debugging purposes; these should be understood as the computational requirements to execute the source code as is.

1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295

Table 1: CPU time and maximum memory required to generate figures

Figure	CPU Time (seconds)	Max Memory (GB)
Figure 3(a)	413.1	< 1
Figure 3(b)	81.9	< 1
Figure 3(c)	40.4	< 1
Figure 4(a) [†]	0.7	< 1
Figure 4(b) [†]	1.1	< 1
Figure 4(c) [†]	80.1	1.9
Figure 5(a) and Figure 6(a)	7,883.0	65.5
Figure 5(b) and Figure 6(b)	2,935.2	17.0
Figure 7(a)-(b) and Figure 8(a)-(b) [‡]	17,531.6	1.6
Figure 7(c)-(d) and Figure 8(c)-(d) [‡]	17,416.3	1.6
Figure 9	527.9	< 1

[†] Figure 4 shows the results with one random seed, but here the CPU Time reports the average over 10 seeds, and Max Memory reports the maximum over 10 seeds.

[‡] Figure 8 could have been obtained by simply reusing and shifting Figure 7, but here the CPU time reports the time to generate the two figures without reuse.

A.5 DETAILS OF SECTION 8

A.5.1 LIMITATIONS

While the proposed approach makes major advancement in the field, it certainly has limitations. Here, we discuss four major limitations of this work as well as interesting directions of research motivated by those limitations.

First, when types are not independent between players, the sufficient condition in Lemma 1 may not be necessary (Proposition 1). This means that the LP may be feasible even when the condition in the lemma is violated, and our results do not provide optimal solutions for those cases. Further research is needed to understand exactly when the sufficient condition is also necessary. It is also important to develop efficient methods for solving the LP when the types are dependent.

Second, our mechanisms guarantee strong budget balance (SBB) and individual rationality (IR) in expectation with respect to the distribution of the players' types, but this does not guarantee that those properties are satisfied *ex post* (for any realization of the types). Although the satisfaction in expectation is often sufficient for risk-neutral decision makers, it is important to let the mediator and the participants aware that they may experience negative utilities even if their expected utilities are nonnegative. It would also be an interesting direction of research to extend the proposed approach towards achieving these properties *ex post*.

Finally, our experiments have considered environments with up to 128 players, each with at most 16 types. Although these are substantially larger than the environment studied in Osogami et al. (2023), they certainly do not cover the scale needed for all applications. While (10-100 times) larger environments could be handled with improved implementation and greater computational resources, essentially new ideas would be needed for substantially (over 10^3 times) larger environments or continuous type space. It would be an interesting direction of research to identify and exploit structures of particular environments for designing scalable approaches of mechanism design for those environments.

A.5.2 SOCIETAL IMPACTS

We expect that the proposed approach has several positive impacts on trading networks in particular and the society in general. In particular, the proposed approach enables mechanisms that can maximize the efficiency of a trading network and minimize the fees that the participants need to pay to the mediator. Also, the DSIC guaranteed by the proposed approach would make it more difficult for malicious participants to manipulate the outcome of a trading network.

1296 On the other hand, the proposed approach might have negative impacts depending on where and
1297 how it is applied. For example, although the proposed approach guarantees individual rationality,
1298 some of the participants might get less benefits from the mechanism designed with our approach
1299 than other participants. This can happen, because maximizing the social welfare does not mean that
1300 all the participants are treated fairly. Before applying the mechanisms designed with the proposed
1301 approach, it is thus recommended to assess whether such fairness needs to be considered and to take
1302 any actions that mitigate the bias if needed.

1304 B PROOFS

1306 B.1 PROOFS ON THE LEMMAS AND COROLLARIES IN SECTION 5

1308 *Proof of Lemma 1.* We will show that

$$1309 \quad h_n(t_{-n}) = \eta_n := \min_{t_n \in \mathcal{T}_n} \{\mathbb{E}[w^*(t) | t_n] - \theta(t_n)\} \quad \forall t_{-n} \in \mathcal{T}_{-n} \quad (37)$$

1311 is a feasible solution when (13) holds.

1313 The θ -IR (11) is satisfied with (37), because for any $n \in \mathcal{N}$ we have

$$1314 \quad \mathbb{E}[h_n(t_{-n}) | t_n] = \eta_n \quad (38)$$

$$1315 \quad = \min_{t'_n \in \mathcal{T}_n} \{\mathbb{E}[w^*(t') | t'_n] - \theta(t'_n)\} \quad (39)$$

$$1316 \quad \leq \mathbb{E}[w^*(t) | t_n] - \theta(t_n). \quad (40)$$

1319 The ρ -WBB (12) is satisfied with (37), because

$$1321 \quad \sum_{n \in \mathcal{N}} \mathbb{E}[h_n(t_{-n})] = \sum_{n \in \mathcal{N}} \eta_n \quad (41)$$

$$1322 \quad = \sum_{n \in \mathcal{N}} \min_{t_n \in \mathcal{T}_n} \{\mathbb{E}[w^*(t) | t_n] - \theta(t_n)\} \quad (42)$$

$$1323 \quad \geq (N - 1) \mathbb{E}[w^*(t)] + \rho, \quad (43)$$

1324 where the inequality follows from (13). \square

1329 *Proof of Lemma 2.* By independence, (11) is reduced to

$$1331 \quad \mathbb{E}[w^*(t) | t_n] - \mathbb{E}[h_n(t_{-n})] \geq \theta(t_n) \quad \forall t_n \in \mathcal{T}_n, \forall n \in \mathcal{N} \quad (44)$$

$$1332 \quad \iff \min_{t_n \in \mathcal{T}_n} \{\mathbb{E}[w^*(t) | t_n] - \theta(t_n)\} \geq \mathbb{E}[h_n(t_{-n})] \quad \forall n \in \mathcal{N}. \quad (45)$$

1334 This together with (12) establishes the necessity of

$$1336 \quad (N - 1) \mathbb{E}[w^*(t)] + \rho \leq \sum_{n \in \mathcal{N}} \mathbb{E}[h_n(t_{-n})] \quad (46)$$

$$1337 \quad \leq \sum_{n \in \mathcal{N}} \min_{t_n \in \mathcal{T}_n} \{\mathbb{E}[w^*(t) | t_n] - \theta(t_n)\}. \quad (47)$$

1341 \square

1343 *Proof of Lemma 3.* We first rewrite the LP (10)-(12) in the following equivalent form:

$$1344 \quad \min_h \quad \sum_{n \in \mathcal{N}} \sum_{t_n \in \mathcal{T}_n} \mathbb{P}[t_n] \mathbb{E}[h_n(t_{-n}) | t_n] \quad (48)$$

$$1345 \quad \text{s.t.} \quad \mathbb{E}[w^*(t) | t_n] - \mathbb{E}[h_n(t_{-n}) | t_n] \geq \theta(t_n) \quad \forall t_n \in \mathcal{T}_n, \forall n \in \mathcal{N} \quad (49)$$

$$1346 \quad \sum_{n \in \mathcal{N}} \sum_{t_n \in \mathcal{T}_n} \mathbb{P}[t_n] \mathbb{E}[h_n(t_{-n}) | t_n] - (N - 1) \mathbb{E}[w^*(t)] \geq \rho. \quad (50)$$

1350 Then it can be easily observed that the optimal objective value must be equal to $(N-1)\mathbb{E}[w^*(t)] + \rho$
 1351 (i.e., when equality holds in (50)), since changing h in a way it decreases the value of $\mathbb{E}[h_n(t_{-n}) |$
 1352 $t_n]$ only makes (49) more satisfiable².

1353 Hence, to prove that any $h \in \mathcal{H}$ is an optimal solution, it suffices to show that (50) is satisfied with
 1354 equality and (49) is satisfied with any $h \in \mathcal{H}$. When $h \in \mathcal{H}$, we have, for any t_n , that

$$1355 \mathbb{E}[h_n(t_{-n}) | t_n] = \eta_n \quad (51)$$

$$1357 \leq \min_{t'_n \in \mathcal{T}_n} \{\mathbb{E}[w^*(t) | t'_n] - \theta(t'_n)\} \quad (52)$$

$$1358 \leq \mathbb{E}[w^*(t) | t_n] - \theta(t_n), \quad (53)$$

1360 where the first inequality follows from (16). We also have

$$1361 \sum_{n \in \mathcal{N}} \sum_{t_n \in \mathcal{T}_n} \mathbb{P}[t_n] \mathbb{E}[h_n(t_{-n}) | t_n] = \sum_{n \in \mathcal{N}} \sum_{t_n \in \mathcal{T}_n} \mathbb{P}[t_n] \eta_n \quad (54)$$

$$1362 = \sum_{n \in \mathcal{N}} \eta_n \quad (55)$$

$$1363 = (N-1)\mathbb{E}[w^*(t)] + \rho, \quad (56)$$

1364 where the last equality follows from (15) and (17).

1365 Finally, when (13) holds, \mathcal{H} is nonempty, because the following η_n satisfies the conditions (15)-(17):

$$1366 \eta_n = \min_{t_n \in \mathcal{T}_n} \{\mathbb{E}[w^*(t) | t_n] - \theta(t_n)\} - \delta \quad (57)$$

1367 where

$$1368 \delta := \frac{1}{N} \left(\sum_{n \in \mathcal{N}} \min_{t_n \in \mathcal{T}_n} \{\mathbb{E}[w^*(t) | t_n] - \theta(t_n)\} - (N-1)\mathbb{E}[w^*(t)] - \rho \right). \quad (58)$$

1369 Notice that $\delta \geq 0$ follows from (13). □

1370 *Proof of Proposition 1.* We construct an example that satisfies (11)-(12) but violates (13). Let $\mathcal{N} =$
 1371 $\{1, 2\}$; $\mathcal{T}_n = \mathcal{T} := \{1, 2\}, \forall n \in \mathcal{N}$; $\rho = 0$; $\theta(m) = 0, \forall m \in \mathcal{T}$. We assume that the types are
 1372 completely dependent (namely, $t_1 = t_2$ surely) and let p be the probability that $t_1 = t_2 = 1$ (hence,
 1373 $t_1 = t_2 = 2$ with probability $1 - p$).

1374 For this example, we rewrite (11)-(12) and (13) by using $x_m := w^*((m, m))$ and $y_{nm} := h_n(m)$
 1375 for $m \in \mathcal{T}$ and $n \in \mathcal{N}$. Notice that, for any $m \in \mathcal{T}$ and $n \in \mathcal{N}$, we have

$$1376 \mathbb{E}[w^*(t) | t_n = m] = x_m \quad (59)$$

$$1377 \mathbb{E}[h_n(t_{-n}) | t_n = m] = y_{nm}, \quad (60)$$

1378 since types are completely dependent. Hence, (11) is reduced to

$$1379 x_m - y_{nm} \geq 0 \quad \forall m \in \mathcal{T}, \forall n \in \mathcal{N} \quad (61)$$

1380 and (12) is reduced to

$$1381 p(y_{11} + y_{21} - x_1) + (1-p)(y_{12} + y_{22} - x_2) \geq 0. \quad (62)$$

1382 On the other hand, (13) is reduced to

$$1383 2 \min\{x_1, x_2\} \geq p x_1 + (1-p) x_2. \quad (63)$$

1384 Consider the case where $x_m > 0, \forall m \in \mathcal{T}$. In this case, (61)-(62) suggest that (11)-(12) are satisfied
 1385 as long as y_{nm} satisfies

$$1386 \frac{x_m}{2} \leq y_{nm} \leq x_m \quad \forall m \in \mathcal{T}, \forall n \in \mathcal{N}, \quad (64)$$

1387 whether (63) is satisfied or not. Indeed, (64) can be met even if (63) is violated, for example when
 1388 $p = \frac{1}{2}, x_1 = 1, x_2 = 4, y_{nm} = \frac{2}{3}x_m, \forall m \in \mathcal{T}, \forall n \in \mathcal{N}$; this serves as an desired example,
 1389 concluding the proof. □

1390 ²This implies that ρ -SBB is satisfied whenever ρ -WBB is satisfied

1404 *Proof of Corollary 2.* The sufficiency follows from Lemma 3. The necessity follows in exactly the
 1405 same way as the proof of Lemma 2. \square

1407
 1408 *Proof of Corollary 3.* By (15), we have

$$1409 \sum_{n \in \mathcal{N}} \eta_n - (N-1) \mathbb{E}[w^*(t)] - \rho$$

$$1410 = \sum_{n \in \mathcal{N}} \min_{t_n \in \mathcal{T}_n} \{\mathbb{E}[w^*(t) | t_n] - \theta(t_n)\} - \sum_{n \in \mathcal{N}} \delta_n - (N-1) \mathbb{E}[w^*(t)] - \rho \quad (65)$$

$$1411 = 0, \quad (66)$$

1412 where the last equality follows from (17). Hence, (12) is satisfied with equality. \square

1413
 1414 *Proof of Corollary 4.* This corollary can be proved analogously to Corollary 3. \square

1415
 1416 *Proof of Corollary 5.* By (15), for any $t_n \in \mathcal{T}_n$ and $n \in \mathcal{N}$, we have

$$1417 \mathbb{E}[w^*(t) | t_n] - \eta_n - \theta(t_n)$$

$$1418 = \mathbb{E}[w^*(t) | t_n] - \theta(t_n) - \left(\min_{t'_n \in \mathcal{T}_n} \{\mathbb{E}[w^*(t) | t'_n] - \theta(t'_n)\} \right) + \delta_n, \quad (67)$$

1419 which is nonnegative by (16), and hence (11) holds. \square

1420 B.2 PROOFS OF THE LEMMAS, THEOREM, AND PROPOSITION IN SECTION 6

1421
 1422 *Proof of Lemma 4.* Since the sample complexity of PAC-BAI in Step 1 is M and Step 3 pulls an
 1423 arm at most m^* times, the sample complexity of Algorithm 3 is at most $M + m^*$. Hence, it remains
 1424 to prove (23).

1425 Recall that \hat{I} is a random variable representing the index of the best arm returned by Algorithm 2
 1426 (PAC-BAI). Then we have the following bound:

$$1427 \Pr \left(|\hat{\mu}_{\hat{I}} - \mu_{\star}| > \frac{3}{2} \varepsilon \right)$$

$$1428 = \Pr \left(\hat{\mu}_{\hat{I}} > \mu_{\star} + \frac{3}{2} \varepsilon \right) + \Pr \left(\hat{\mu}_{\hat{I}} < \mu_{\star} - \frac{3}{2} \varepsilon \right) \quad (68)$$

$$1429 \leq \Pr \left(\hat{\mu}_{\hat{I}} > \mu_{\hat{I}} + \frac{3}{2} \varepsilon \right)$$

$$1430 + \Pr \left(\left\{ \hat{\mu}_{\hat{I}} < \mu_{\star} - \frac{3}{2} \varepsilon \right\} \cap \{ \mu_{\hat{I}} < \mu_{\star} - \varepsilon \} \right) + \Pr \left(\left\{ \hat{\mu}_{\hat{I}} < \mu_{\star} - \frac{3}{2} \varepsilon \right\} \cap \{ \mu_{\hat{I}} \geq \mu_{\star} - \varepsilon \} \right) \quad (69)$$

$$1431 \leq \Pr \left(\hat{\mu}_{\hat{I}} > \mu_{\hat{I}} + \frac{3}{2} \varepsilon \right) + \Pr \left(\mu_{\hat{I}} < \mu_{\star} - \varepsilon \right) + \Pr \left(\left\{ \hat{\mu}_{\hat{I}} < \mu_{\star} - \frac{3}{2} \varepsilon \right\} \cap \{ \mu_{\hat{I}} \geq \mu_{\star} - \varepsilon \} \right) \quad (70)$$

$$1432 \leq \Pr \left(\hat{\mu}_{\hat{I}} > \mu_{\hat{I}} + \frac{3}{2} \varepsilon \right) + \Pr \left(\mu_{\hat{I}} < \mu_{\star} - \varepsilon \right) + \Pr \left(\hat{\mu}_{\hat{I}} < \mu_{\hat{I}} - \frac{1}{2} \varepsilon \right) \quad (71)$$

$$1433 \leq \Pr \left(\hat{\mu}_{\hat{I}} > \mu_{\hat{I}} + \frac{3}{2} \varepsilon \right) + \delta + \Pr \left(\hat{\mu}_{\hat{I}} < \mu_{\hat{I}} - \frac{1}{2} \varepsilon \right), \quad (72)$$

1434 where the last inequality follows from PAC(ε, δ) of BAI.

1435 Since $\hat{\mu}_{\hat{I}}$ is the average of $M' := \max\{M_{\hat{I}}, m^*\}$ samples from arm \hat{I} , where $M_{\hat{I}}$ is the number of
 1436 times Algorithm 2 (PAC-BAI) has pulled arm \hat{I} , we have $M' \geq m^*$. Hence, applying Hoeffding's

inequality to the last term of (72), we obtain

$$\Pr\left(\hat{\mu}_{\hat{I}} < \mu_{\hat{I}} - \frac{1}{2}\varepsilon\right) = \sum_{n \geq m^*, k \in [1, K]} \Pr\left(\hat{\mu}_k < \mu_k - \frac{1}{2}\varepsilon \mid M' = n, \hat{I} = k\right) \Pr(M' = n, \hat{I} = k) \quad (73)$$

$$\leq \sum_{n \geq m^*, k \in [1, K]} \exp\left(-2 \left(\frac{1}{2}\varepsilon\right)^2 n\right) \Pr(M' = n, \hat{I} = k) \quad (74)$$

$$\leq \exp\left(-2 \left(\frac{1}{2}\varepsilon\right)^2 m^*\right), \quad (75)$$

where the first inequality is obtained by applying Hoeffding's inequality to the sample mean $\hat{\mu}_k$ of n independent random variables having support in $[0, 1]$, and the second inequality follows from $n \geq m^*$. We can also show the following inequality in an analogous manner:

$$\Pr\left(\hat{\mu}_{\hat{I}} > \mu_{\hat{I}} + \frac{3}{2}\varepsilon\right) \leq \exp\left(-2 \left(\frac{3}{2}\varepsilon\right)^2 m^*\right). \quad (76)$$

By applying (75)-(76) to (72), we finally establish the bound to be shown:

$$\Pr\left(|\hat{\mu}_{\hat{I}} - \mu_{\star}| > \frac{3}{2}\varepsilon\right) \leq \delta + \exp\left(-2 \left(\frac{3}{2}\varepsilon\right)^2 m^*\right) + \exp\left(-2 \left(\frac{1}{2}\varepsilon\right)^2 m^*\right) \quad (77)$$

$$\leq \delta + \exp\left(-2 \left(\frac{3}{2}\varepsilon\right)^2 \frac{2}{\varepsilon^2} \log \frac{1.22}{\delta}\right) + \exp\left(-2 \left(\frac{1}{2}\varepsilon\right)^2 \frac{2}{\varepsilon^2} \log \frac{1.22}{\delta}\right) \quad (78)$$

by the definition of m^*

$$= \delta + \left(\frac{\delta}{1.22}\right)^9 + \frac{\delta}{1.22} \quad (79)$$

$$\leq \left(1 + \frac{1}{1.22^9} + \frac{1}{1.22}\right) \delta \quad (80)$$

$$\leq 2\delta. \quad (81)$$

□

Proof of Lemma 7. Although the lemma is stated in Even-Dar et al. (2002) with reference to Chernoff (1972), this specific lemma is neither stated nor proved explicitly in Chernoff (1972). For completeness, here, we prove the lemma following the general methodology provided in Chernoff (1972). Specifically, we derive the expected sample size required by the sequential probability-ratio test (SPRT; Section 10 of Chernoff (1972)), whose optimality (Theorem 12.1 of Chernoff (1972)) will then establish the lemma.

Consider two hypotheses, θ_1 and θ_2 , for the probability distribution \mathbb{P} of a random variable X , which takes either the value of 1 or -1 , where

$$\mathbb{P}(X = 1 \mid \theta_1) = \frac{1 + \varepsilon}{2} \quad (82)$$

$$\mathbb{P}(X = 1 \mid \theta_2) = \frac{1 - \varepsilon}{2} \quad (83)$$

Consider the SPRT procedure that takes i.i.d. samples, X_1, X_2, \dots, X_N , from \mathbb{P} until the stopping time N when

$$\lambda_N := \prod_{n=1}^N \frac{\mathbb{P}(X_n \mid \theta_1)}{\mathbb{P}(X_n \mid \theta_2)} = \prod_{n=1}^N \left(\frac{\varepsilon + 1}{\varepsilon - 1}\right)^{X_n} \quad (84)$$

hits either $A \in \mathbb{R}$ or $1/A$. When λ_N hits A , we identify θ_1 as the correct hypothesis. When λ_N hits $1/A$, we identify θ_2 as the correct hypothesis.

1512 Let

$$1513 S_N := \log \lambda_N = \sum_{n=1}^N X_n \log \frac{1+\varepsilon}{1-\varepsilon}. \quad (85)$$

1516 Since N is a stopping time, by Wald's lemma, we have

$$1518 \mathbb{E}[S_N | \theta_1] = \mathbb{E}[N | \theta_1] \mathbb{E}[X | \theta_1] \log \frac{1+\varepsilon}{1-\varepsilon} \quad (86)$$

$$1520 = \mathbb{E}[N | \theta_1] \varepsilon \log \frac{1+\varepsilon}{1-\varepsilon} \quad (87)$$

$$1522 \mathbb{E}[S_N | \theta_2] = -\mathbb{E}[N | \theta_2] \varepsilon \log \frac{1+\varepsilon}{1-\varepsilon}. \quad (88)$$

1524 Let δ be the probability of making the error in identifying the correct hypothesis. Then we must have

$$1527 \mathbb{E}[S_N | \theta_1] = (1-\delta) \log A + \delta \log(1/A) \quad (89)$$

$$1528 \mathbb{E}[S_N | \theta_2] = \delta \log A + (1-\delta) \log(1/A). \quad (90)$$

1529 By (86)-(90), we have

$$1531 \mathbb{E}[N | \theta_1] = \mathbb{E}[N | \theta_2] = \frac{1-2\delta}{\varepsilon \log \frac{1+\varepsilon}{1-\varepsilon}} \log A. \quad (91)$$

1534 Now, notice that S_N hits $\log A$ when we have

$$1536 |\{n : X_n = 1\}| - |\{n : X_n = -1\}| \geq \frac{\log A}{\log \frac{1+\varepsilon}{1-\varepsilon}} \quad (92)$$

1538 for the first time and hits $-\log A$ when we have

$$1540 |\{n : X_n = -1\}| - |\{n : X_n = 1\}| \geq \frac{\log A}{\log \frac{1+\varepsilon}{1-\varepsilon}} \quad (93)$$

1542 for the first time. Hence, by the gambler's ruin probability, we have

$$1544 \delta = \frac{1 - \left(\frac{1+\varepsilon}{1-\varepsilon}\right)^{\frac{\log A}{\log \frac{1+\varepsilon}{1-\varepsilon}}}}{1 - \left(\frac{1+\varepsilon}{1-\varepsilon}\right)^{2 \frac{\log A}{\log \frac{1+\varepsilon}{1-\varepsilon}}}} = \frac{1}{1 + \left(\frac{1+\varepsilon}{1-\varepsilon}\right)^{\frac{\log A}{\log \frac{1+\varepsilon}{1-\varepsilon}}}}, \quad (94)$$

1548 which implies

$$1550 A = \frac{1-\delta}{\delta}. \quad (95)$$

1552 Plugging the last expression into (91), we obtain

$$1554 \mathbb{E}[N | \theta_1] = \mathbb{E}[N | \theta_2] = \frac{1-2\delta}{\varepsilon \log \frac{1+\varepsilon}{1-\varepsilon}} \log \frac{1-\delta}{\delta} = O\left(\frac{1}{\varepsilon^2} \log \frac{1}{\delta}\right). \quad (96)$$

1557 □

1558 *Proof of Lemma 5.* We will construct an algorithm that correctly identifies the mean α of B with probability at least δ using at most M/K samples of B in expectation given the access to an $(\varepsilon/2, \delta/2)$ -PAC BME with sample complexity M . The algorithm first draws i^+ and i^- independently from a uniform distribution over $[1, K]$.

1563 Consider two environments of K -armed bandit, \mathcal{E}^+ and \mathcal{E}^- , where every arm gives the reward according to the Bernoulli distribution with mean $\alpha^- = (1-\varepsilon)/2$ except that the reward of arm i^+ in \mathcal{E}^+ has the same distribution as B and the reward of arm i^- in \mathcal{E}^- has the same distribution as $1-B$. Note that the algorithm can simulate the reward with the known mean α^- using the

algorithm's internal random number generators. Only when the algorithm pulls arm i^+ of \mathcal{E}^+ or arm i^- of \mathcal{E}^- , it uses the sample of B , which contributes to the sample complexity of the algorithm.

The algorithm then runs two copies of the $(\varepsilon/2, \delta/2)$ -PAC BME with sample complexity M in parallel: one referred to as BME^+ is run on \mathcal{T}^+ , and the other referred to as BME^- is run on \mathcal{T}^- . At each step, let M^+ be the number of samples BME^+ has taken from arm i^+ by that step and M^- be the corresponding number BME^- has taken from arm i^- . If $M^+ < M^-$, the algorithm lets BME^+ pull an arm; otherwise, the algorithm lets BME^- pull an arm. Therefore, $|M^+ - M^-| \leq 1$ at any step.

This process is continued until one of BME^+ and BME^- terminates and returns an estimate $\hat{\mu}$ of the best mean. If BME^+ terminates first, then the algorithm determines that $\alpha = \alpha^-$ if $\hat{\mu} < 1/2$ and that $\alpha = \alpha^+$ otherwise. If BME^- terminates first, then the algorithm determines that $\alpha = \alpha^+$ if $\hat{\mu} < 1/2$ and that $\alpha = \alpha^-$ otherwise. Due to the $(\varepsilon/2, \delta/2)$ -PAC property of BME^+ and BME^- , the algorithm correctly identifies the mean of B with probability at least $1 - \delta$. Formally, if $\alpha = \alpha^+$, we have

$$\Pr\left(\hat{\mu} < \frac{1}{2}\right) = \Pr\left(\hat{\mu} < \frac{1}{2} \mid \text{BME}^+ \text{ terminates first}\right) \Pr(\text{BME}^+ \text{ terminates first}) \\ + \Pr\left(\hat{\mu} < \frac{1}{2} \mid \text{BME}^- \text{ terminates first}\right) \Pr(\text{BME}^- \text{ terminates first}) \quad (97)$$

$$\leq \frac{\delta}{2} + \frac{\delta}{2} \quad (98)$$

$$= \delta. \quad (99)$$

Analogously, $\Pr(\hat{\mu} < \frac{1}{2}) \leq \delta$ can be shown if $\alpha = \alpha^-$.

What remains to prove is the sample complexity of the algorithm. Recall that each of BME^+ and BME^- pulls arms at most M times before it terminates due to their sample complexity. Notice that the arms in \mathcal{E}^- are indistinguishable when $\alpha = \alpha^-$, and the arms in \mathcal{E}^+ are indistinguishable when $\alpha = \alpha^+$. Therefore, at least one of BME^+ and BME^- is run on the environment where the arms are indistinguishable. Since i^+ and i^- are sampled uniformly at random from $[1, K]$, BME (either BME^+ or BME^-) would take at most M/n samples from B in expectation if the arms are indistinguishable. Since $|M^+ - M^-| \leq 1$, we establish that the sample complexity of the algorithm is $O(M/n)$ in expectation. \square

Proof of Lemma 6. Note that the PAC estimators can give independent estimates, $\tilde{\kappa}_n(\theta)$ for $n \in \mathcal{N}$ and $\tilde{\lambda}$, such that

$$\Pr(|\tilde{\kappa}_n(\theta) - \kappa_n(\theta)| \leq \varepsilon') \geq 1 - \delta' \quad \forall n \in \mathcal{N} \quad (100)$$

$$\Pr(|\tilde{\lambda}(\theta) - \lambda(\theta)| \leq \varepsilon'') \geq 1 - \delta', \quad (101)$$

which imply

$$\Pr\left(|\tilde{\lambda}(\theta) - \lambda(\theta)| \leq \varepsilon'', |\tilde{\kappa}_n(\theta) - \kappa_n(\theta)| \leq \varepsilon', \forall n \in \mathcal{N}\right) \geq (1 - \delta')^{N+1}. \quad (102)$$

Hence, with probability at least $(1 - \delta')^{N+1}$, the expected utility of player n given its type t_n (recall (8)) is

$$\mathbb{E}[w^*(t) \mid t_n] - (\tilde{\kappa}_n(\theta) - \tilde{d}_n) \geq \mathbb{E}[w^*(t) \mid t_n] - \kappa_n(\theta) - \varepsilon' + \varepsilon''' \quad (103)$$

$$\geq \theta(t_n) - (\varepsilon' - \varepsilon'''), \quad (104)$$

where the first inequality follows from the PAC bound on $\tilde{\kappa}_n(\theta)$ and the definition of \tilde{d}_n , and the last inequality follows from the original guarantee when $\kappa_n(\theta)$ is exactly computed, and the expected revenue of the mediator (recall (9)) is

$$\sum_{n \in \mathcal{N}} (\tilde{\kappa}_n(\theta) - \tilde{d}_n) - (N - 1) \mathbb{E}[w^*(t)] = (N - 1) (\tilde{\lambda} + \varepsilon''') - (N - 1) \mathbb{E}[w^*(t)] \quad (105)$$

$$\geq \rho + (N - 1) (\varepsilon''' - \varepsilon''), \quad (106)$$

where the equality follows from the definition of \tilde{d}_n , and the inequality follows from the PAC bound on $\tilde{\lambda}$. \square

1620 *Proof of Theorem 2.* With the choice of the parameters in the theorem, we have

$$1621 \theta - (\varepsilon' - \varepsilon''') = \theta \quad (107)$$

$$1622 (\rho - (N - 1)(\varepsilon'' - \varepsilon''')) = \rho \quad (108)$$

$$1623 (1 - \delta')^{1/(N+1)} = 1 - \delta. \quad (109)$$

1624 Hence, Lemma 6 guarantees that the constant pivot rule $h_n(t_{-n}) = \tilde{\kappa}_n(\theta) - \tilde{d}_n$ satisfies DSIC, DE,
 1625 θ -IR, and ρ -WBB with probability $1 - \delta$. \square

1626
 1627
 1628
 1629
 1630
 1631
 1632
 1633
 1634
 1635 *Proof of Proposition 3.* The constant pivot rule in Theorem 2 can be learned with N independent
 1636 runs of an (ε, δ') -PAC BME and a single run of an (ε, δ') -PAC estimator for an expectation, whose
 1637 overall sample complexity is $O(N(K/\varepsilon^2) \log(1/\delta'))$. The proposition can then be established by
 1638 substituting $\delta' = 1 - (1 - \delta)^{1/(N+1)}$. \square

1639 B.3 PROOF OF THE PROPOSITIONS IN SECTION 7

1640
 1641
 1642
 1643 *Proof of Proposition 5.* Let $\mu_\star := \max_{k \in [1, K]} \mu_k$ be the best mean, $\hat{\mu}$ be the best mean estimated by
 1644 Algorithm 4, and $\hat{\mu}_k^{(t)}$ be the average of the first t samples from arm k . Let $\alpha_t := \sqrt{\frac{1}{2t} \log\left(\frac{\pi^2 K t^2}{3\delta}\right)}$.
 1645 Then we have

$$1646 \Pr(|\hat{\mu} - \mu_\star| \leq \varepsilon)$$

$$1647 \geq \Pr(\text{At every iteration, the error in the estimated mean is less than } \alpha_t \text{ for any arm in } \mathcal{R}.^3)$$

$$1648 \quad (110)$$

$$1649 = \Pr\left(\bigcap_{t=1}^{\infty} \bigcap_{k \in [1, K]} \{|\hat{\mu}_k^{(t)} - \mu_k| < \alpha_t\}\right)$$

$$1650 \quad (111)$$

$$1651 = 1 - \Pr\left(\bigcup_{t=1}^{\infty} \bigcup_{k \in [1, K]} \{|\hat{\mu}_k^{(t)} - \mu_k| \geq \alpha_t\}\right)$$

$$1652 \quad (112)$$

$$1653 \geq 1 - \sum_{t=1}^{\infty} \sum_{k \in [1, K]} \Pr\left(\{|\hat{\mu}_k^{(t)} - \mu_k| \geq \alpha_t\}\right) \quad \text{by union bound}$$

$$1654 \quad (113)$$

$$1655 \geq 1 - 2K \sum_{t=1}^{\infty} \exp(-2\alpha_t^2 t) \quad \text{by Hoeffding's inequality}$$

$$1656 \quad (114)$$

$$1657 = 1 - \delta \frac{6}{\pi^2} \sum_{t=1}^{\infty} \frac{1}{t^2} \quad \text{by definition of } \alpha_t$$

$$1658 \quad (115)$$

$$1659 = 1 - \delta. \quad (116)$$

1660 \square

1661
 1662
 1663
 1664
 1665
 1666
 1667
 1668
 1669
 1670
 1671
 1672
 1673 ³This condition suffices because it ensures that the best arms always remain in \mathcal{R} .

1674 *Proof of Proposition 6.* Let $\alpha_t := \sqrt{\log\left(\frac{\pi^2 K t^2}{6\delta}\right) / (2t)}$. Let \mathcal{B} be the set of the (strictly) best arms.

1675 Let $\hat{\mu}_k^{(t)}$ be the average of the first t samples from arm k . Then we have

1677 $\Pr(\text{Algorithm 5 selects an } \varepsilon\text{-best arm.})$
 1678 $\geq \Pr(\text{At every iteration, all arms in } \mathcal{B} \text{ remain in } \mathcal{R} \text{ and any arm in } \mathcal{R} \text{ is } 2\alpha_t\text{-best.}^4)$ (117)

1680 $\geq \Pr\left(\bigcap_{t=1}^{\infty} \bigcap_{k \in \mathcal{B}} \{\hat{\mu}_k^{(t)} > \mu_k - \alpha_t\} \bigcap_{\ell \notin \mathcal{B}} \{\hat{\mu}_\ell^{(t)} < \mu_\ell + \alpha_t\}\right)$ (118)

1682 $= 1 - \Pr\left(\bigcup_{i=1}^{\infty} \bigcup_{k \in \mathcal{B}} \{\hat{\mu}_k^{(i)} \leq \mu_k - \alpha_i\} \bigcup_{\ell \notin \mathcal{B}} \{\hat{\mu}_\ell^{(i)} \geq \mu_\ell + \alpha_i\}\right)$ (119)

1684 $\geq 1 - \sum_{i=1}^{\infty} \left(\sum_{k \in \mathcal{B}} \Pr(\hat{\mu}_k^{(i)} \leq \mu_k - \alpha_i) + \sum_{\ell \notin \mathcal{B}} \Pr(\hat{\mu}_\ell^{(i)} \geq \mu_\ell + \alpha_i) \right)$ by union bound (120)

1686 $\geq 1 - \sum_{i=1}^{\infty} K \exp(-2\alpha_i^2 i)$ by Hoeffding's inequality (121)

1688 $= 1 - \delta \frac{6}{\pi^2} \sum_{i=1}^{\infty} \frac{1}{i^2}$ by the definition of α_t (122)

1690 $= 1 - \delta.$ (123)

1691 □

1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727

⁴This condition suffices because the algorithm stops either when $|\mathcal{R}| = 1$ or when $\alpha_t \leq \varepsilon/2$, which implies that only ε -best arms are in \mathcal{R} when the algorithm stops.