From Bayesian to Crowdsourced Bayesian Auctions: When Everything is Known by Somebody

Jing Chen Bo Li Yingkai Li
Department of Computer Science, Stony Brook University
Stony Brook, NY 11794
{jingchen, boli2, yingkli}@cs.stonybrook.edu

April 6, 2016

Abstract

In Bayesian auction design, it is assumed that the distributions of the players' values are common knowledge to the seller and the players —the *common prior* assumption. A lot of effort has been made in the literature trying to remove this assumption when, for example, the seller can observe independent samples from the distributions before the auction begins.

In this work, we focus on the problem when the seller has no knowledge at all and the players have knowledge about each other (like long-time competitors in the same market). We formalize the intuitive idea that "nobody knows all, but everything is known by somebody," and design mechanisms that generate good revenue by crowdsourcing the players' individual knowledge. We emphasize that the seller does not know the players' distributions or true values, nor does he know which player knows what. We consider two information models.

- First, *everything is known by somebody*: that is, for each player i and item j, there exists at least k other players who know the distribution of i's value for item j, where k can be any number between 1 and $n-1$, with n being the number of players. We design mechanisms for auctions of unit-demand and auctions of additive valuations, two widely studied multi-parameter auctions. Our mechanisms are constant approximations to the optimal Bayesian mechanisms for generating revenue even when $k = 1$, and their revenue converges to that of the best known Bayesian mechanisms when k gets larger (but still much smaller than n).
- Second, *everybody is known by somebody*: that is, for each player i, there exists at least k other players who know the distribution of i's valuation function. For *any* combinatorial auction, our mechanism uses any Bayesian mechanism as a blackbox and achieves a τ_k-approximation to the latter, where $\tau_k = \frac{k}{(k+1)^{\frac{k+1}{k}}}$. Note that $\tau_1 = 1/4$ and $\tau_k \to 1$ when k gets larger.

Our work tries to provide a general approach for understanding the effect of partial knowledge on the revenue of Bayesian auctions and for designing crowdsourced Bayesian mechanisms.

Keywords: game theory, mechanism design, distributed knowledge, multi-parameter settings

1 Introduction

We consider auctions for generating revenue, by selling m heterogenous items to n potential buyers—the players. Each player has a valuation function describing how much he values each subset of the items. In single-parameter settings [40, 1], including single-good auctions [41] and single-minded auctions [36], a player i's valuation function is specified by a single value; while in auctions of unit demand [15] and auctions of additive valuations [29], two widely studied multi-parameter settings, a player's valuation function is specified by m values, one for each item. The valuation function of a player is his private information.

In Bayesian mechanism design, the valuation profile of the players is drawn from an underlying distribution, which is assumed to be common knowledge to the seller and the players—the *common prior* assumption. Following [40], Bayesian mechanism design has been extremely flourishing in the past several decades and led to many auction mechanisms that are optimal or approximately optimal in generating revenue; see, e.g., [21, 42, 38, 15, 17, 12, 47] as a highly incomplete list.

However, as pointed out by [45], the common prior assumption is very demanding and is unlikely to hold in real-world scenarios. Where does the seller get information about the distribution? How much revenue can he guarantee if he does not know it? A lot of effort has been made trying to answer these questions and remove the common prior assumption from Bayesian mechanism design. These works come from both the economics literature (see, e.g., [9, 2, 8]) and the computer science literature (see, e.g., [18, 31, 20]). We will discuss them more carefully in Section 1.2.

How crucial is the common prior assumption to Bayesian mechanisms? This is a question we would like to understand. In this paper, we consider a much weaker assumption, *crowdsourced Bayesian*. Roughly speaking, the seller has no knowledge about the players at all and the players have knowledge about each other, like long-time competitors in the same market. It is possible that no player knows how the whole valuation profile is distributed, but some players individually know some other players' distributions. We emphasize that the seller does not know the players' distributions or true valuations, nor does he know which player knows what.

To our best knowledge, the only work in the literature that is close to our model is [4], where each player individually knows the whole distribution or a refinement of it. In essence, [4] assumes "every player knows all", while we assume "everything is known by somebody but maybe nobody knows all." Interestingly, this intuitive idea behind our model has long been considered by philosophers. In [34], the author discussed a world where "everything ... might be known by somebody, yet not everything by the same knower."

In this paper, we formalize this intuition for the crowdsourced Bayesian model and construct approximately optimal auction mechanisms based on it. Our model is defined in Section 2. Below we summarize our main results.

1.1 Main Results

To keep the settings simple and highlight our techniques, we start by assuming the players will not *bluff*. That is, a player will not report anything about a distribution that he does not know. In the last part of the paper, we remove this assumption from all our results. Indeed, this assumption serves as a way for us to focus on the key problems in designing crowdsourced Bayesian auctions. We design mechanisms that are *2-step dominant strategy truthful* (2-DST) [4]. That is, fixing any knowledge about the others that player i may report, it is dominant for i to report his true values; and given that all players report their true values, it is dominant for each player to report his true knowledge about others. Due to lack of space, most proofs are provided in the Appendix.

Partial information: everything is known by somebody. First, we consider multi-parameter auctions where a player i's value for an item j, v_{ij}, is drawn from a distribution $\mathcal{D}_{ij}$, and all values are drawn independently. Instead of a common prior, we assume that *each* $\mathcal{D}_{ij}$ *is known by at least* k *other players.* When $k = 1$, this is the weakest assumption in the crowdsourced Bayesian model.

Representing the players' knowledge as a graph, we have the following theorem for single-good auctions, proved in Section 3. Here OPT is the revenue of the optimal *dominant strategy truthful* (DST) Bayesian mechanism[1] and ϵ is a positive value that can be arbitrarily small.

Theorem 1. (restated) *When the knowledge graph is 2-connected, there is a 2-DST crowdsourced Bayesian mechanism for single-good auctions, whose expected revenue is at least* $(1 - \frac{1}{n})OPT - \epsilon$.

We say this mechanism is a $(1 - \frac{1}{n})$-approximation. For unit-demand auctions we have the following (see Section 4.1), where $\tau_k = \frac{k}{(k+1)^{\frac{k+1}{k}}}$. Note that $\tau_1 = \frac{1}{4}$ and $\tau_k \to 1$ as k gets larger.

Theorem 2. (restated) $\forall k \in [n-1]$*, when each distribution is known by at least* k *players, there is a 2-DST crowdsourced Bayesian mechanism for unit-demand auctions with revenue* $\geq \frac{\tau_k}{6} \cdot OPT - \epsilon$.

Furthermore, for additive auctions we have the following, proved in Section 4.2. Here OPT_B is the revenue of the optimal *Bayesian incentive compatible* (BIC) mechanism[2] and $OPT_B \geq OPT$.

Theorem 3. (restated) $\forall k \in [n-1]$*, when each distribution is known by at least* k *players, there is a 2-DST crowdsourced mechanism for additive auctions with revenue* $\geq \max\{\frac{1}{24}, \frac{\tau_k}{12}\}OPT_B - \epsilon$.

Player-wise information: every player is known by somebody. Next, we consider combinatorial auctions in general. In the worst case, a player's valuation function is specified by 2^m values, one for each subset. In this setting, for each player i, his valuation function v_i is drawn from a distribution $\mathcal{D}_i$. Different players' valuation functions are drawn independently, but a player's own values for different items and/or subsets of items can be arbitrarily correlated. Note that such auctions include multi-parameter auctions above as special cases.

To deal with this complex setting, we assume that each $\mathcal{D}_i$ is known by at least k other players, hence the name *player-wise information.* We show a blackbox reduction from any Bayesian mechanism to a crowdsourced Bayesian mechanism and we have the following theorem, see Section 5.

Theorem 4. (restated) $\forall k \in [n-1]$*, when each* $\mathcal{D}_i$ *is known by at least* k *players, given any BIC mechanism* $\mathcal{M}_B$ *that is a* σ*-approximation to* OPT *(respectively,* OPT_B*), there is a BIC crowdsourced Bayesian mechanism* $\mathcal{M}_{CSB}$ *with revenue* $\geq \tau_k \sigma OPT - \epsilon$ *(respectively,* $\geq \tau_k \sigma OPT_B - \epsilon$*). Moreover, if* $\mathcal{M}_B$ *is DST then* $\mathcal{M}_{CSB}$ *is 2-DST.*

Removing the no-bluff assumption. In Section 6, we show how to remove the no-bluff assumption used by our mechanisms. *All* results obtained so far can be carried through, with the mechanisms being BIC instead of 2-DST. Accordingly, we do not explicitly state the theorem here.

Techniques and future directions. In this paper, we model the players' knowledge as *knowledge graphs* and show for the first time the relationships between the graph structures and the approximation ratios of crowdsourced Bayesian mechanisms. In the future, it would be interesting to precisely characterize such relationships.

Scoring rules are widely used in social choices and prediction markets; see, e.g., [10, 26, 46, 6, 39]. Given the crowdsourced feature of our model, it is natural to use scoring rules in our mechanisms,

[1] A Bayesian mechanism is DST if it is dominant for each player to report his true valuation function.

[2] A Bayesian mechanism is BIC if it is a Bayesian equilibrium for all players to always report their true valuations.

although we elicit information from multiple experts who themselves are potential buyers in the auctions. We combine scoring rules with random sampling, and the approximation ratios in our main results are obtained by decomposing multi-parameter auctions into single-parameter ones similar to [16] and [47], and by patching smaller randomly-sampled auctions into larger ones to boost the lower bound for their revenue. Interestingly, we also use scoring rules to remove the no-bluff assumption. In a nutshell, *not only a player is paid to report things that he knows, he is also paid for "staying quiet about things that he does not know and letting the experts talk."*

Finally, in some sense, our results demonstrate that the common prior assumption is without much loss of generality if one is willing to give up some portion of the revenue, and this portion shrinks smoothly as the amount of knowledge increases: that is, as k grows from 1 to $n-1$. As k grows larger (but can still be much smaller than n), our results converge to the best known Bayesian mechanisms. Since a crowdsourced Bayesian mechanism implies a Bayesian mechanism, an important direction is to quantify the performance gaps between the two for every k, which will precisely answer how much one is losing when there is less and less knowledge in the system.

1.2 Additional Related Work

Bayesian auction design. In his seminal work [40], Myerson introduced the first optimal Bayesian mechanism for single-good auctions with independent values, which applies to many other single-parameter settings [1]. In the past several decades, there has been a huge literature on designing optimal or approximately optimal Bayesian mechanisms for generating revenue, such as [41, 42, 15, 32, 25, 14]. The readers may refer to [30] for a thorough introduction.

Mechanisms that are (approximately) optimal for multi-parameter settings have been constructed recently. In [11, 12], the authors characterize the structures of optimal Bayesian mechanisms for various multi-parameter settings. For unit-demand auctions, [17, 16, 35] provide constant approximations to the optimal mechanism. For auctions with additive valuations, [29, 37, 47] provide various constant or logarithmic approximations under different conditions. [13] provides a duality-based unified approach for designing Bayesian mechanisms and obtains better approximation ratios for unit-demand auctions and additive auctions. Moreover, [43] studied revenue mechanisms with a single buyer and sub-additive valuations.

Removing the common prior assumption. Following [45], a lot of effort has been made in the literature trying to remove the common prior assumption from Bayesian mechanism design [7, 44, 27]. In particular, in *prior-free mechanisms* [31, 22], the distribution is unknown and the designer learns about it from the values of a group of randomly selected players. In [20, 24, 33, 23], the seller can observe independent samples from the distribution before the auction begins. In [18, 19], the players have arbitrary possibilistic (rather than probabilistic) beliefs/belief hierarchies about each other and the seller generates good revenue by leveraging such beliefs. Different from a player's *knowledge* which is always correct, a player's *belief* may be wrong and different players' beliefs may be inconsistent. In *robust mechanism design* [8, 2, 9], the players have arbitrary probabilistic belief hierarchies and the studies have characterized necessary and sufficient conditions for a social correspondence to be robustly implementable.

Finally, in [4], the authors consider crowdsourced Bayesian auctions where each player knows the distribution of the whole valuation profile or a refinement about it. In some sense, their setting is a special case of ours, where a distribution about player i is known by all the other $n-1$ players. As shown in the technical part of this paper, when each distribution is known by k players and when k gets larger, the approximation ratios of our mechanisms approach theirs, but our mechanisms work for arbitrary k and arbitrary structure of the knowledge graphs.

2 The Crowdsourced Bayesian Model

For unit-demand auctions and additive auctions, each v_{ij}, player i's value for item j, is independently drawn from $\mathcal{D}_{ij}$. For general combinatorial auctions, each v_i, player i's valuation function, is independently drawn from $\mathcal{D}_i$. All of our settings are downward closed [32]. The players have quasi-linear utilities and are risk neutral. We use Brier's scoring rule [10] in our mechanisms, denoted by BSR and shifted up so that its range is $[0, 2]$. Readers may refer to Appendix A for a more detailed introduction to Bayesian mechanism design. Below we introduce our model.

Knowledge graphs. It is illustrative to define the players' knowledge as a directed graph. In single-parameter settings, the *knowledge graph* has n nodes, one for each player, and there is an edge from i to j if and only if player i knows the distribution $\mathcal{D}_j$ of player j's value.[3] In multi-parameter settings, we distinguish two cases: *partial information*, where for each player i and item j, there exists a player i' who knows $\mathcal{D}_{ij}$; and *player-wise information*, where, for each player i, there exists a player i' who knows the whole distribution $\mathcal{D}_i$ of i's valuation. Accordingly, the partial information setting corresponds to m knowledge graphs, one for each item, and each player may only know part of $\mathcal{D}_i$; while the player-wise information setting corresponds to a single knowledge graph.

A knowledge graph is *connected* if there is a directed path from any player i to any other player i'; and it is *2-connected* if it remains connected after removing any single player and its adjacent edges. Intuitively, being connected means that for each i and i', "i knows somebody who knows somebody ... who knows i'"; and being 2-connected means that there does not exist a "crucial" player who appears in all paths from one part of the graph to the other part.

For any $k \in [n-1]$, a knowledge graph is *k-bounded* if each node has in-degree at least k: a player's distribution is known by at least k other players. In partial information settings, different distributions may be known by different groups of players. A knowledge graph being connected or 2-connected respectively implies that it is 1-bounded or 2-bounded, but not vice versa.

Notice that a partial information setting with 1-bounded knowledge graphs is a very weak assumption to make about the players' knowledge, and in fact the *weakest* assumption in the crowdsourced Bayesian model. On the other hand, the common prior assumption implies that the knowledge graphs (whether of partial information or player-wise information) are complete graphs, or $(n-1)$-bounded, which is the strongest assumption.

Crowdsourced Bayesian mechanisms. Different from Bayesian mechanisms, a crowdsourced Bayesian mechanism does not have the distribution $\mathcal{D}$ as input. Instead, it asks each player i to report a valuation function b_i and a *knowledge* K_i, where K_i is a distribution for the valuation subprofile v_{-i}. K_i may contain the symbol "$\perp$" at some places, indicating that i does not know the distributions of the corresponding values. The mechanism maps each strategy profile to an allocation and a price profile, which may be randomized. Bayesian Incentive Compatibility are defined in the same way for crowdsourced Bayesian mechanisms as for Bayesian mechanisms.

The no-bluff assumption. As mentioned before, we make the *no-bluff assumption* about the players in Sections 3-5; and we show how to remove it in Section 6. A player i in a crowdsourced Bayesian mechanism is *no-bluff* if, for any knowledge graph G and player j such that $(i, j) \notin G$, and for any strategy (b_i, K_i) of i, i reports $\perp$ for the corresponding distribution of j. Notice that this assumption applies to both the partial information setting and the player-wise information setting. Also notice that, for a player j with $(i, j) \in G$, i may report any distribution about j, including $\perp$.

[3] We could have defined the players' knowledge using the standard notion of Harsanyi [28] and Aumann [3]: roughly speaking, the state space consists of all possible distributions of the whole valuation profile and player i knows $\mathcal{D}_j$ if he is in an information set where all distributions have the j-th component equal to $\mathcal{D}_j$. However, the knowledge graph is a more succinct representation and is enough for the purpose of this paper.

3 Crowdsourced Single-Good Auctions

As a warm up, in this section, we construct a crowdsourced Bayesian mechanism for single-good auctions with 2-connected knowledge graphs. Given the set N of players, the set M of items and the distribution $\mathcal{D}$ of the value profile, let $\hat{\mathcal{I}} = (N, M, \mathcal{D})$ denote the Bayesian auction instance and $\mathcal{I} = (N, M, \mathcal{D}, G)$ a corresponding crowdsourced Bayesian instance, where G is a knowledge graph. To distinguish whether a mechanism $\mathcal{M}$ is a Bayesian or crowdsourced Bayesian mechanism, we explicitly write $\mathcal{M}(\hat{\mathcal{I}})$ or $\mathcal{M}(\mathcal{I})$. Notice that the latter does not mean $\mathcal{M}$ has $\mathcal{D}$ or G as its input.

It is well known that Myerson's mechanism [40] achieves optimal revenue among all Bayesian mechanisms. It maps the players' values to the *virtual value* space and runs the second-price mechanism with reserve price 0 in the latter space. When the distributions are irregular, the "ironed" virtual values are used instead. For each player j, value b_j and distribution $\mathcal{D}_j$, we use $\phi_j(b_j; \mathcal{D}_j)$ to denote the (ironed) virtual value of b_j, assuming that j's value is distributed according to $\mathcal{D}_j$. Note that ϕ_j is monotone in b_j. The inverse function $\phi_j^{-1}(\cdot\,; \mathcal{D}_j)$ is defined respectively. Our Crowdsourced Myerson Mechanism $\mathcal{M}_{CSM}$ is defined in Mechanism 1. It uses Myerson's mechanism in a non-blackbox way and takes as input a positive number ϵ which can be arbitrarily small.

Mechanism 1. $\mathcal{M}_{CSM}$

1: Simultaneously, each player i reports a value b_i and a knowledge $K_i = (\mathcal{D}^i_j)_{j \neq i}$.
2: Randomly choose a player a, set $S = \{j \mid \mathcal{D}^a_j \neq \perp\}$, $N' = N \setminus (\{a\} \cup S)$, and $\mathcal{D}'_j = \mathcal{D}^a_j \ \forall j \in S$.
3: If $S = \emptyset$, the item is unsold, the mechanism sets price $p_i = 0\ \forall i \in N$ and goes to Step 14.
4: Set $i^* = \arg\max_{j \in S} \phi_j(b_j; \mathcal{D}'_j)$, with ties broken lexicographically.
5: **while** $N' \neq \emptyset$ **do**
6:    Set $S' = \{j \mid j \in N',\ \exists i' \in S \setminus \{i^*\} \text{ s.t. } \mathcal{D}^{i'}_j \neq \perp\}$.
7:    If $S' = \emptyset$ then go to Step 12.
8:    For each $j \in S'$, set $\mathcal{D}'_j = \mathcal{D}^{i'}_j$, where i' is the first player in $S \setminus \{i^*\}$ with $\mathcal{D}^{i'}_j \neq \perp$.
9:    Set $S = \{i^*\} \cup S'$ and $N' = N' \setminus S'$.
10:    Set $i^* = \arg\max_{j \in S} \phi_j(b_j; \mathcal{D}'_j)$, with ties broken lexicographically.
11: **end while**
12: Set $\phi_{second} = \max_{j \in N \setminus (\{a, i^*\} \cup N')} \phi_j(b_j; \mathcal{D}'_j)$ and the price $p_i = 0$ for each player i.
13: If $\phi_{i^*}(b_{i^*}; \mathcal{D}'_{i^*}) < 0$ then the item is unsold; otherwise, the item is sold to player i^* and $p_{i^*} = \phi^{-1}_{i^*}(\max\{\phi_{second}, 0\}; \mathcal{D}'_{i^*})$.
14: Finally, for each pair of players i and i' such that $\mathcal{D}^i_{i'} \neq \perp$, set $p_i = p_i - \frac{\epsilon}{2n^2} \cdot BSR(\mathcal{D}^i_{i'}, b_{i'})$.

To help understand $\mathcal{M}_{CSM}$, we illustrate in Figure 1 of Appendix B the sets of players involved in the first round. Under the no-bluff assumption, Theorem 1 below shows that $\mathcal{M}_{CSM}$ is almost as powerful as Myerson's mechanism, where $Rev(\mathcal{M})$ denotes the revenue of a mechanism $\mathcal{M}$.

Theorem 1. *For any single-good auction instances $\hat{\mathcal{I}} = (N, M, \mathcal{D})$ and $\mathcal{I} = (N, M, \mathcal{D}, G)$ where G is 2-connected, $\mathcal{M}_{CSM}$ is 2-DST and $\mathbb{E}_{v \sim \mathcal{D}} Rev(\mathcal{M}_{CSM}(\mathcal{I})) \geq (1 - \frac{1}{n}) OPT(\hat{\mathcal{I}}) - \epsilon$.*

The following lemma is proved in Appendix C.

Lemma 1. *$\mathcal{M}_{CSM}$ is 2-DST.*

Proof of Theorem 1. It remains to show $\mathbb{E}_{v \sim \mathcal{D}} Rev(\mathcal{M}_{CSM}(\mathcal{I})) \geq (1 - \frac{1}{n}) OPT(\hat{\mathcal{I}}) - \epsilon$. Arbitrarily fix the player a chosen by the mechanism. Notice that N' is the set of players $i \in N \setminus \{a\}$ such that $\mathcal{D}'_i$ is not defined. We show that $N' = \emptyset$ at the end of the mechanism.

Indeed, since G is 2-connected, the out-degree of a in G is at least 2: otherwise, either a cannot reach any other node in G, or this becomes the case after removing the unique node j with $(a, j) \in G$, contradicting 2-connectedness. Since a reports his true knowledge K_a, we have $|S| \geq 2$ in Step 3 and the mechanism does not go to Step 14 directly from there. Moreover, at the beginning of each round, we have $|S| \geq 2$ and thus $S \setminus \{i^*\} \neq \emptyset$: otherwise $S' = \emptyset$ in the previous round, and the mechanism would not have reached this round.

Assume the mechanism finally reaches a round r where $N' \neq \emptyset$ at the beginning but $S' = \emptyset$ in Step 6. Since all players report their true knowledge, by the definition of S' we have that, in graph G, all neighbors of $S \setminus \{i^*\}$ are in $N \setminus N'$. Furthermore, for any player $i \in (N \setminus N') \setminus S$, we have that in graph G, all neighbors of i are also in $N \setminus N'$: indeed, i has been moved from N' to S and then dropped from S (except player a, whose neighbors are in $N \setminus N'$ by definition); and when i is dropped from S, all his neighbors in N' are moved to S. Accordingly, all the out-going edges of $N \setminus N'$ are from player i^* and G becomes disconnected by removing i^*, a contradiction. Thus, $S' \neq \emptyset$ in all rounds and $N' = \emptyset$ in the end, as we wanted to show.

Because all players report their true values and true knowledge, we have $\mathcal{D}'_{-a} = \mathcal{D}_{-a}$ and $\phi_i(v_i; \mathcal{D}'_i) = \phi_i(v_i; \mathcal{D}_i)$ for all $i \neq a$. Letting $\hat{\mathcal{I}}_a = (N \setminus \{a\}, M, \mathcal{D}_{-a})$, we have

$$\mathbb{E}_{v \sim \mathcal{D}}[Rev(\mathcal{M}_{CSM}(\mathcal{I}))|a] = OPT(\hat{\mathcal{I}}_a) - \epsilon. \tag{1}$$

To see why Equation 1 is true, note that by construction, in each round the mechanism keeps the player with the highest virtual value in S. Thus, player i^* at the end has the highest virtual value in $N \setminus \{a\}$ and ϕ_{second} is the second highest virtual value in $N \setminus \{a\}$. Accordingly, the outcome produced in Step 13 is the same as that of Myerson's mechanism on $\hat{\mathcal{I}}_a$, so is the revenue. Since Brier's scoring rule is bounded in $[0, 2]$, the reward each player i gets in the last step is no more than $\frac{\epsilon}{n}$ and the total reward given to the players is no more than ϵ. Therefore Equation 1 holds.

Finally, for each player i, letting $P_i(OPT(\hat{\mathcal{I}}))$ be the expected price paid by player i in Myerson's mechanism under $\hat{\mathcal{I}}$, we have $OPT(\hat{\mathcal{I}}) = \sum_{i \in N} P_i(OPT(\hat{\mathcal{I}}))$. Consider the following Bayesian mechanism $\mathcal{M}'$ on $\hat{\mathcal{I}}_a$: it runs Myerson's mechanism on $\hat{\mathcal{I}}$ and then projects the outcome to players $N \setminus \{a\}$. It is easy to see that $\mathcal{M}'$ is DST and its expected revenue is $\sum_{i \neq a} P_i(OPT(\hat{\mathcal{I}}))$. As $\mathcal{M}'$ cannot general more revenue than the optimal mechanism for $\hat{\mathcal{I}}_a$, we have

$$OPT(\hat{\mathcal{I}}_a) \geq \mathbb{E}_{\mathcal{D}_{-i}} Rev(\mathcal{M}'(\hat{\mathcal{I}}_a)) = \sum_{i \neq a} P_i(OPT(\hat{\mathcal{I}})). \tag{2}$$

Combining Equations 1 and 2, we have

$$\mathbb{E}_{v \sim \mathcal{D}} Rev(\mathcal{M}_{CSM}(\mathcal{I})) = \sum_{a \in N} \frac{1}{n} \mathbb{E}_{v \sim \mathcal{D}}[Rev(\mathcal{M}_{CSM}(\mathcal{I}))|a] = \sum_{a \in N} \frac{1}{n} \left(OPT(\hat{\mathcal{I}}_a) - \epsilon \right)$$

$$\geq \sum_{a \in N} \frac{1}{n} \left(\sum_{i \neq a} P_i(OPT(\hat{\mathcal{I}})) - \epsilon \right) = \left(\sum_{i \in N} \frac{n-1}{n} P_i(OPT(\hat{\mathcal{I}})) \right) - \epsilon = (1 - \frac{1}{n}) OPT(\hat{\mathcal{I}}) - \epsilon,$$

and Theorem 1 holds. □

Remark. Since ϵ can be arbitrarily small and an ϵ additive discount is always used in our mechanisms, for simplicity, in the remaining part of the paper we will say, for example, that mechanism $\mathcal{M}_{CSM}$ is a $(1 - \frac{1}{n})$-approximation. Following [4], no crowdsourced Bayesian mechanism can be a $(\frac{1}{2} + \delta)$-approximation for any positive constant δ when $n = 2$, thus our result is tight.

Finally, notice that in the last step of $\mathcal{M}_{CSM}$, we reward a player i for every distribution $\mathcal{D}^i_{i'} \neq \perp$ reported by him. A slightly more complex construction is to reward i only if $\mathcal{D}^i_{i'}$ has been used by the mechanism: that is, if $\mathcal{D}'_{i'}$ is defined by $\mathcal{D}^i_{i'}$. By doing so, the reward i gets can be increased to $\frac{\epsilon}{2n} \cdot BSR$ and the total reward is still upper bounded by ϵ.

4 Partial Information: Everything Is Known by Somebody

In this section, we consider multi-parameter auctions with partial information. Let $\hat{\mathcal{I}} = (N, M, \mathcal{D})$ be a Bayesian auction instance and $\mathcal{I} = (N, M, \mathcal{D}, G)$ a corresponding crowdsourced Bayesian instance. Here $G = (G_1, \ldots, G_m)$ is a vector of knowledge graphs, one for each item; and $\mathcal{D} = \times_{i \in N, j \in M} \mathcal{D}_{ij}$. For each $k \in [n-1]$, we say that G is k-bounded if all the G_j's are k-bounded. Moreover, let $\tau_k = \frac{k}{(k+1)^{\frac{k+1}{k}}}$. The value τ_k will serve as an important parameter in our analysis.

4.1 Unit-demand Auctions

For unit-demand auctions, [35] has constructed a DST Bayesian mechanism $\mathcal{M}_{UD}$ which is a 6-approximation. In this subsection, we construct a crowdsourced Bayesian mechanism $\mathcal{M}_{CSUD}$ using random sampling and using $\mathcal{M}_{UD}$ as a blackbox. Intuitively, the mechanism asks one group of players to report their knowledge about another group, and uses $\mathcal{M}_{UD}$ to generate revenue from the latter. However, the analysis is far from a direct application of the Bayesian mechanism's approximation ratio, as will become clear in the proof of our theorem. The mechanism takes k and ϵ as inputs and is shown in Mechanism 2. We have the following theorem, proved in Appendix D.

Mechanism 2. $\mathcal{M}_{CSUD}$

1: Each player i reports a valuation $b_i = (b_{ij})_{j \in M}$ and a knowledge $K_i = (\mathcal{D}^i_{i'j})_{i' \neq i, j \in M}$.
2: Randomly partition the players into two sets, N_1 and N_2, where each player is independently put in N_1 w.p. q and N_2 w.p. $1-q$. The value q will be decided in the analysis.
3: Set $N_3 = \emptyset$ and $p_i = 0$ for each player i.
4: **for** players $i \in N_1$ lexicographically **do**
5: For each player $i' \in N_2$ and item $j \in M$, if $\mathcal{D}'_{i'j}$ has not been defined yet and $\mathcal{D}^i_{i'j} \neq \bot$, then (1) set $\mathcal{D}'_{i'j} = \mathcal{D}^i_{i'j}$ and reward player i using Brier's scoring rule: that is, $p_i = p_i - \frac{\epsilon}{mn} BSR(\mathcal{D}'_{i'j}, b_{i'j})$; and (2) add player i' to N_3.
6: **end for**
7: For each $i \in N_3$ and $j \in M$ such that $\mathcal{D}'_{ij}$ is not defined, set $\mathcal{D}'_{ij}$ to be the distribution which is 0 with probability 1, and set b_{ij} to be 0.
8: Run $\mathcal{M}_{UD}$ on the unit-demand Bayesian auction $(N_3, M, (\mathcal{D}'_{ij})_{i \in N_3, j \in M})$, with the players' values being $(b_{ij})_{i \in N_3, j \in M}$. Let $X = (x'_{ij})_{i \in N_3, j \in M}$ be the resulting allocation where $x'_{ij} \in \{0, 1\}$, and let $P = (p'_i)_{i \in N_3}$ be the prices. Without loss of generality, $x'_{ij} = 0$ if $\mathcal{D}'_{ij}$ is 0 with probability 1.
9: For each player $i \notin N_3$, i gets no item and his price is p_i.
10: For each player $i \in N_3$, i gets item j if $x'_{ij} = 1$, and his price is $p_i = p'_i$.

Theorem 2. *$\forall k \in [n-1]$, unit-demand auction instances $\hat{\mathcal{I}} = (N, M, \mathcal{D})$ and $\mathcal{I} = (N, M, \mathcal{D}, G)$ where G is k-bounded, $\mathcal{M}_{CSUD}$ is 2-DST and $\mathbb{E}_{v \sim \mathcal{D}} Rev(\mathcal{M}_{CSUD}(\mathcal{I})) \geq \frac{\tau_k}{6} \cdot OPT(\hat{\mathcal{I}}) - \epsilon$.*

Remark. When $k = 1$, $\mathcal{M}_{CSUD}$ is a 24-approximation. Since $\tau_k \to 1$ as k gets larger, the approximation ratio converges to that of the Bayesian mechanism, even when $k << n$.

Although we use $\mathcal{M}_{UD}$ as a blackbox, the analysis is "non-blackbox" and it remains open whether there is a blackbox reduction from Bayesian mechanisms to crowdsourced Bayesian mechanisms with partial information. When $k = n-1$, [4] provides such a reduction with a $(1 - \frac{1}{n})$ loss in the approximation ratio. In our mechanism, $\tau_{n-1} = \frac{n-1}{n^{n/(n-1)}} \to 1 - \frac{1}{n}$ when n gets larger. Thus, our revenue essentially matches that of [4], except that our mechanism works for all $k \in [n-1]$.

Finally, by [13], $OPT(\hat{\mathcal{I}}^{CP}) \geq \frac{1}{4} \cdot OPT_B(\hat{\mathcal{I}})$, so we immediately have that mechanism $\mathcal{M}_{CSUD}$ is a $\frac{\tau_k}{24}$-approximation to the optimal BIC mechanism for unit-demand auctions.

4.2 Additive Auctions

In this subsection we focus on auctions with additive valuations. Let $\hat{\mathcal{I}} = (N, M, \mathcal{D})$ be a Bayesian instance where $\mathcal{D} = \times_{i\in N, j\in M}\mathcal{D}_{ij}$, and $\mathcal{I} = (N, M, \mathcal{D}, G)$ be a corresponding crowdsourced Bayesian instance, where $G = (G_1, \ldots, G_m)$ is a vector of knowledge graphs. Some of the ideas here are similar to those for unit-demand auctions, and we have omitted the detailed discussion when this happens.

In [47], the author first gave a 69-approximation DST Bayesian mechanism for this setting, where the seller may choose from two independent mechanisms and, given the instance $\hat{\mathcal{I}}$, he runs the one with higher expected revenue. One of the two mechanisms is the *Individual Second-price* mechanism, where the seller sells each item separately using the second-price mechanism. The other one is called *Bundle VCG* and denoted by $BVCG$ for short. Under $BVCG$, each player i announces $b_i = (b_{ij})_{j\in M}$ and the mechanism computes an *entry fee* $e_i(\mathcal{D}_i, b_{-i})$ for i, which depends on i's distribution and the other players' reported values. For each item j, the potential winner is $i^*(j)$ such that $i^*(j) = \arg\max_i \{b_{ij}\}$, with ties broken lexicographically. The price for item j is $p_j = \max_{i\neq i^*(j)} \{b_{ij}\}$. For each player i, let M_i be the set of items where he is the potential winner. If $\sum_{j\in M_i} b_{ij} \geq \sum_{j\in M_i} p_j + e_i(\mathcal{D}_i, b_{-i})$, player i will get all the items in M_i, with price $p_i = \sum_{j\in M_i} p_j + e_i(\mathcal{D}_i, b_{-i})$. Otherwise, player i will receive no item and pay nothing.

Recently, [13] shows that by replacing the Individual Second-price mechanism with Myerson's mechanism for each item, a better approximation ratio can be achieved. In particular, denoting the latter as mechanism IM (for "Individual Myerson"), [13] shows that $\max\{Rev(IM), Rev(BVCG)\} \geq \frac{OPT_B}{8}$, where the tight-hand side is the expected revenue of the optimal BIC mechanism, which of course is an upper bound for the expected revenue of the optimal DST Bayesian mechanism.

It is worth pointing out that in Bayesian settings, the seller knows $\mathcal{D}$, so he can compute the expected revenue of both mechanisms and run the better one. In crowdsourced Bayesian settings, however, the seller relies on the players' reported knowledge. If the mechanism computes the expected revenue of the two mechanisms based on the reported knowledge and runs the better one, then the players may not be truthful. Indeed, a player's utility in the better mechanism may be very low, while his utility in the other mechanism may be very high, giving him incentives to lie about his knowledge. One could use the same random sampling idea as in unit-demand auctions, letting one group of players report their knowledge about the other group, comparing the two mechanisms on the latter and running the better one. However, a difficulty is that the $BVCG$ mechanism's revenue may be much smaller when, for each player, only a subset of items whose distributions are reported can be used for computing the entry fee, as it sells items in bundles.

Instead, we construct a crowdsourced Bayesian mechanism for each of the two mechanisms IM and $BVCG$, denoted by $\mathcal{M}_{CSIM}$ and $\mathcal{M}_{CSBVCG}$ respectively, and let the seller choose between them randomly. To further improve the approximation ratio, we construct a crowdsourced Bayesian mechanism based on the 1-Lookahead mechanism of [42], denoted by $\mathcal{M}_{CS1LA}$, and use it to better approximate mechanism IM. Combining these gadgets together, we have a crowdsourced Bayesion mechanism for additive auctions, denoted by $\mathcal{M}_{CSA}$. Below we only state our theorem, and the mechanisms will be defined during the analysis, provided in Appendix D.

Theorem 3. *$\forall k \in [n-1]$, additive auction instances $\hat{\mathcal{I}} = (N, M, \mathcal{D})$ and $\mathcal{I} = (N, M, \mathcal{D}, G)$ where G is k-bounded, $\mathcal{M}_{CSA}$ is 2-DST and $\mathbb{E}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{CSA}(\mathcal{I})) \geq \max\{\frac{1}{24}, \frac{\tau_k}{12}\} \cdot OPT_B(\hat{\mathcal{I}}) - \epsilon$.*

When the knowledge graphs are 2-connected, we can use mechanism $\mathcal{M}_{CSM}$ from Section 3 to improve the approximation ratio of $\mathcal{M}_{CSA}$: instead of using $\mathcal{M}_{CS1LA}$ and $\mathcal{M}_{CSIM}$, we use $\mathcal{M}_{CSM}$ for each item j. We have the following corollary.

Corollary 1. *When each knowledge graph G_j is 2-connected, the redefined mechanism $\mathcal{M}_{CSA}$ is 2-DST and $\mathop{\mathbb{E}}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{CSA}(\mathcal{I})) \geq \frac{1}{12}(1-\frac{1}{n})OPT_B(\hat{\mathcal{I}}) - \epsilon$.*

5 Player-wise Information: Everybody Is Known by Somebody

In this section, we consider general combinatorial auctions in crowdsourced Bayesian settings with player-wise information. Here a Bayesian instance is denoted by $\hat{\mathcal{I}} = (N, M, \mathcal{D} = \times_{i \in N} \mathcal{D}_i)$, where $\mathcal{D}_i$ is the distribution of i's valuation function, and i's values for different items or subsets of items can be arbitrarily correlated. In particular, our results here apply to all unit-demand auctions and additive auctions.[4] A corresponding crowdsourced Bayesian instance is denoted by $\mathcal{I} = (N, M, \mathcal{D}, G)$, where G is a single knowledge graph. Recall that an edge $(i, i') \in G$ means that player i knows $\mathcal{D}_{i'}$.

Given any Bayesian mechanism $\mathcal{M}_B$ as a blackbox, our crowdsourced Bayesian mechanism $\mathcal{M}_{CSB}$ uses random sampling to elicit the players' knowledge and runs $\mathcal{M}_B$ with the collected information. Our mechanism is defined in Mechanism 3, with details that appeared in previous mechanisms omitted. We have the following theorem, proved in Appendix E.

Mechanism 3. $\mathcal{M}_{CSB}$

1: Each player i reports a valuation function b_i and a knowledge $K_i = (\mathcal{D}^i_{i'})_{i' \neq i}$.
2: Randomly partition the players into two sets, N_1 and N_2, where each player is independently put in N_1 with probability q and N_2 with probability $1-q$, with $q = 1 - (k+1)^{-\frac{1}{k}}$.
3: Let N_3 be the set of players in N_2 whose distributions are reported by some players in N_1, and let $\mathcal{D}'_{N_3}$ be the vector of reported distributions.
4: Reward players in N_1 using Brier's scoring rule, with total reward $R \leq \epsilon$.
5: Run $\mathcal{M}_B$ on the Bayesian instance $\hat{\mathcal{I}}_{N_3} = (N_3, M, \mathcal{D}'_{N_3})$ and the valuation functions b_{N_3}; and use the resulting allocation and prices to sell to players in N_3.

Theorem 4. *For any $k \in [n-1]$, for any auction instances $\hat{\mathcal{I}} = (N, M, \mathcal{D})$ and $\mathcal{I} = (N, M, \mathcal{D}, G)$ where G is k-bounded, if $\mathcal{M}_B$ is DST then $\mathcal{M}_{CSB}$ is 2-DST; and if $\mathcal{M}_B$ is BIC then $\mathcal{M}_{CSB}$ is BIC. Moreover, if $\mathcal{M}_B$ is a σ-approximation to OPT (respectively, OPT_B), then $\mathcal{M}_{CSB}$ is a $\tau_k\sigma$-approximation to OPT (respectively, OPT_B), up to an additive ϵ discount.*

Remark. As before, the approximation ratio of $\mathcal{M}_{CSB}$ tends to match that of the Bayesian mechanism when k gets larger (and is $\sigma/4$ when $k = 1$). When $k = 1$, similar to [4], we have that for any number of players n, no crowdsourced 2-DST mechanism can do better than a 2-approximation; and a similar argument applies to crowdsourced BIC mechanisms.

6 Removing the No-Bluff Assumption

In this section we remove the no-bluff assumption from crowdsourced Bayesian auctions. Without this assumption, player i may report a distribution for another player i' even if $(i, i') \notin G$. However, if there exists a third player $\hat{i}$ who knows i''s distribution $\mathcal{D}_{i'}$, and if player i is also rewarded for player $\hat{i}$'s report, then it is intuitive that i has no incentive to bluff about i'. That is, *not only a player is paid for reporting the distributions he knows, but also he is paid for keeping quiet about the distributions he does not know and letting the experts speak.* Surely, reporting $\mathcal{D}_{i'}$ will maximize i's expected reward, but i does not have the information to decide what $\mathcal{D}_{i'}$ is.[5] Therefore, as long as reporting $\perp$ gives i the same utility as the unknown strategy of reporting $\mathcal{D}_{i'}$, and as long as reporting any distribution other than $\mathcal{D}_{i'}$ gives i a strictly smaller utility (which can be guaranteed by a proper scoring rule), player i will report $\perp$ about i'.

Taking our mechanisms for unit-demand auctions and additive auctions in Section 4 as examples, we can change the reward steps into the following:

[4] Note that in unit-demand auctions and additive auctions studied in the literature and in previous sections, a player i's values for different subsets of items are correlated, but his values for different items are independent.

[5] Using standard notions, player i's information set contains at least two different distributions for i'.

- For each player i and item j, let R_{ij} be the set of players who did not report $\perp$ about i's value for j. Randomly select a player i' from R_{ij} and compute $r_{ij} = BSR(\mathcal{D}^{i'}_{ij}, b_{ij})$, where $\mathcal{D}^{i'}_{ij}$ is the distribution reported by i' and b_{ij} is i's reported value for j. Reward all players other than i using r_{ij}, scaled properly so that the total reward given by the mechanism is at most ϵ.

Other parts of the mechanisms remain the same. It is easy to see that, for any k-bounded crowdsourced Bayesian instance with $k \geq 1$, if all players except i report their true values and true knowledge, then player i's best strategy is to tell the truth about his own. That is, the resulting mechanisms are BIC instead of 2-DST. The mechanisms in Sections 3 and 5 can also be changed in the same way, except that the distributions are for different players rather than player-item pairs. We have the following theorem, whose proof is almost the same as previous ones and thus omitted.

Theorem 5. *For any crowdsourced 2-DST Bayesian mechanism $\mathcal{M}$ in Sections 3, 4 and 5, there exists a crowdsourced BIC mechanism $\mathcal{M}'$ with the same expected revenue in the same auction settings, without making the no-bluff assumption.*

7 Further Extension and Discussions

When Not Everything Is Known. If the knowledge graphs are not even 1-bounded, that is, if for some player i and item j, no player knows the distribution of i's value for j, then it is clear that no crowdsourced Bayesian mechanism can achieve a bounded approximation to the optimal Bayesian mechanism. Indeed, it is possible that player i's value for j is much higher than all other values in the system and is the only source where revenue comes from —for example, all other values are 0 with probability 1. Then the instance is essentially a single-player single-good auction. Although the optimal Bayesian mechanism can find the optimal reserve price based on the distribution, a crowdsourced Bayesian mechanism can only set a fixed reserve price for all distributions as the seller has no knowledge about the true distribution, and the approximation ratio is unbounded.

Accordingly, when *not everything is known*, that is, when $k = 0$ in our model, a meaningful benchmark is the expected revenue of the optimal Bayesian mechanism on players and items for whom the distributions are actually known in the crowdsourced setting. More precisely, letting $\mathcal{I} = (N, \mathcal{D}, G)$ be a single-good crowdsourced auction instance and $N' = \{i \mid \exists i' \text{ s.t. } (i', i) \in G\}$, the benchmark is $OPT(N', \mathcal{D}_{N'})$. For unit-demand auction and additive auction instances $\mathcal{I} = (N, M, \mathcal{D}, G)$, let $\mathcal{D}' = \times_{i \in N, j \in M} \mathcal{D}'_{ij}$ where $\mathcal{D}'_{ij} = \mathcal{D}_{ij}$ if there exists player i' such that $(i', i) \in G_j$, and $\mathcal{D}'_{ij}$ is constant 0 otherwise. Then the benchmark is $OPT(N, M, \mathcal{D}')$. The benchmarks for crowdsourced auctions with player-wise information are defined similarly.

With some effort, it can be seen that when $k = 0$, our mechanisms in Sections 4 and 5 can be changed so as to achieve constant approximations to the benchmarks above, under the no-bluff assumption. We elaborate it in Appendix F. When the players may indeed bluff, however, a player may not report $\perp$ about a distribution he does not know, because he may still get some reward in case nobody knows that distribution. It is an interesting open problem to design crowdsourced Bayesian mechanisms when not everything is known and without making the no-bluff assumption.

Efficiently Reporting the Distributions. In this paper we focus on the revenue performance of crowdsourced Bayesian mechanisms and do not consider the computation or communication complexity. Once the distributions are collected, the computation complexity of our mechanisms is the same as the corresponding Bayesian mechanisms. How to design crowdsourced Bayesian mechanisms that only ask the players to report polynomially many bits about the distributions and can compute the rewards efficiently is another interesting open problem. For example, whether the players can report a discretized approximation of the distributions, or whether it suffices for them to report some particular parameters, such as the expectations or variances.

Appendix

A Additional Preliminaries

In a general combinatorial auction with n players and m items, the set of players is denoted by $N = \{1, \ldots, n\}$ and the set of items by $M = \{1, \ldots, m\}$. Each player i's valuation function v_i maps each subset of items to a non-negative real number and $v_i(\emptyset) = 0$. Bayesian auction design typically considers multi-parameter settings where the players' valuations can be succinctly described. That is, each player i has a value v_{ij} for each item j, and i's value for a subset S of items is $\max_{j \in S} v_{ij}$ in *unit-demand* auctions, or $\sum_{j \in S} v_{ij}$ in *additive* auctions.

The whole valuation profile is drawn from some distribution $\mathcal{D}$ and each player individually knows his own values. It is often assumed that each v_{ij} is independently drawn from a distribution $\mathcal{D}_{ij}$ and $\mathcal{D} = \times_{i,j} \mathcal{D}_{ij}$. In this paper, we also consider settings where each player i's valuation function is independently drawn from a distribution $\mathcal{D}_i$, but his own values for the items may be arbitrarily correlated with each other.

The players' *utilities* are quasi-linear and risk neutral: that is, for each player i, his utility for receiving a subset S at a price p_i is $u_i(S, p_i) = v_i(S) - p_i$, and is the corresponding expectation if S and p_i are random variables.

An *allocation* of an auction specifies a subset of items for each player. The set of feasible allocations an auction can produce is given by some constraints. The most natural constraint is that each item can only be sold once in an allocation. For unit-demand auctions, it may be further required that a player gets at most one item in each allocation. Another widely considered constraint is that the set of allocations is *downward closed* [32]: that is, for any feasible allocation $A = (A_1, \ldots, A_n)$ and any allocation $A' = (A'_1, \ldots, A'_n)$ with $A'_i \subseteq A_i$ for each i, A' is also feasible. Notice that *matroid* constraints [5, 35] are downward closed, but the other direction need not be true. Also notice that downward closed multi-parameter settings can represent any single-parameter setting, such as single-good auctions or single-minded auctions. We always consider downward closed settings in this paper.

A *Bayesian mechanism* asks each player i to report a valuation function b_i and outputs an allocation and a price profile based on $\mathcal{D}$ and $(b_1, \ldots, b_n)$. The *revenue* of a mechanism $\mathcal{M}$, denoted by $Rev(\mathcal{M})$, is the total price paid by the players.

Scoring rules are widely studied in economics and an important ingredient of our mechanisms. A scoring rule is a function f that takes as inputs a distribution $\mathcal{D}$ over a state space Ω and a random sample ω from an underlying true distribution $\mathcal{D}'$ over Ω, and outputs a real number. f is *proper* if $\mathbb{E}_{\omega \sim \mathcal{D}'} f(\mathcal{D}', \omega) \geq \mathbb{E}_{\omega \sim \mathcal{D}'} f(\mathcal{D}, \omega)$ for any $\mathcal{D}$, and *strictly proper* if the inequality is strict for any $\mathcal{D} \neq \mathcal{D}'$. f is *bounded* if there exist constants c_1, c_2 such that $c_1 \leq f(\mathcal{D}, \omega) \leq c_2$ for any $\mathcal{D}$ and ω. In our results, we can use any strictly proper scoring rules that are bounded. For concreteness, we use Brier's scoring rule [10]:

$$BSR(\mathcal{D}, \omega) = 2 - (\sum_{s \in \Omega} (\delta_{\omega,s} - \mathcal{D}(s))^2) = 2\mathcal{D}(\omega) - ||\mathcal{D}||_2^2 + 1,$$

where $\mathcal{D}(s)$ is the probability of s according to $\mathcal{D}$, and $\delta_{\omega,s}$ is the indicator for whether $\omega = s$ or not. Note that $BSR(\mathcal{D}, \omega) \in [0, 2]$ for all $\mathcal{D}$ and ω.[6]

[6]The original version of BSR is bounded by -1 and 1, and we have shifted it up by 2.

B Figures

To help understand mechanism $\mathcal{M}_{CSM}$ in Section 3, we illustrate in Figure 1 the sets of players involved in the first round. The edges in the figure correspond to distributions reported by the players. In each round, the mechanism keeps in S the player with the highest virtual value, drops all other players from S, and adds the players whose distributions are reported for the first time by the dropped ones.

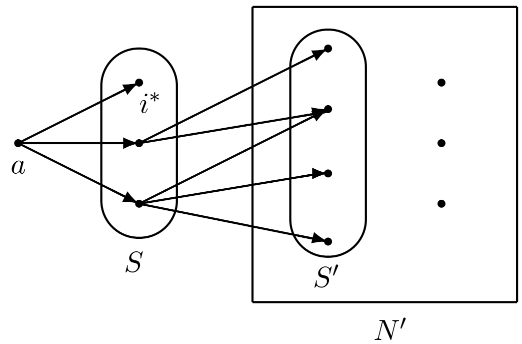

Figure 1: The sets of players involved in the first round of Mechanism $\mathcal{M}_{CSM}$.

C Proofs for Section 3

Lemma 1. (restated) $\mathcal{M}_{CSM}$ *is 2-DST.*

Proof. We start by showing the following claim.

Claim 1. *For any player i, knowledge K_i and true value v_i, fixing the second component of i's strategy to be K_i, it is dominant for i to report v_i.*

Proof. Arbitrarily fix a value b_i and a strategy $s_j = (b_j, K_j)$ for each player $j \neq i$. We need to compare $\mathbb{E}_{\mathcal{M}_{CSM}} u_i(v_i, K_i)$ and $\mathbb{E}_{\mathcal{M}_{CSM}} u_i(b_i, K_i)$, where the expectation is taken over the mechanism's random coins.[7]

First, conditional on $a = i$, player i does not get the item and his reported value is only used in Step 14 to compute the reward of other players who have reported about i's distribution, thus $u_i(v_i, K_i) = u_i(b_i, K_i)$.

Second, we compare the two utilities conditional on $a \neq i$. Notice that when $a \neq i$, whether or not player i's distribution is reported—that is, whether or not $\mathcal{D}'_i$ is defined— only depends on the other players' strategies. Thus, $\mathcal{D}'_i$ is defined under (v_i, K_i) if and only if it is defined under (b_i, K_i).

If $\mathcal{D}'_i$ is not defined, then $i \in N'$ at the end of the mechanism and his utility is equal to the total reward he gets from Brier's scoring rule in Step 14, which solely depends on K_i and b_{-i}. Therefore $u_i(v_i, K_i) = u_i(b_i, K_i)$ again.

If $\mathcal{D}'_i$ is defined, then it is defined in the same round under both (v_i, K_i) and (b_i, K_i), which we refer to as round r. Also, $\mathcal{D}'_i$ is the same in both cases and the mechanism's execution is the same till this round. Notice that

[7]Strictly speaking, each s_j should depend on v_j and the expectation should also be taken over $\mathcal{D}_{-i}$. However, reporting v_i is dominant for player i no matter what v_{-i} is, thus there is no need to consider $\mathcal{D}_{-i}$.

(1) $\phi_i(\cdot; \mathcal{D}'_i)$ is monotone in its input;

(2) i gets the item if and only if $i = i^*$ in all rounds ℓ with $\ell \geq r$ and his virtual value is at least 0; and

(3) when i gets the item, K_i is never used by the mechanism and thus does not affect the execution of any round ℓ with $\ell \geq r$.

Accordingly, the mechanism is *monotone* in player i's reported value: if i gets the item by reporting some value, then he still gets it by reporting a higher value. Moreover, when i gets the item, his price in Step 13 is the *threshold* payment. Following standard characterizations of single-parameter DST mechanisms, we have that, if $\mathcal{D}'_i$ is defined then it is dominant for player i to report v_i: that is, $u_i(v_i, K_i) \geq u_i(b_i, K_i)$ for all s_{-j} such that $\mathcal{D}'_i$ is defined, and the inequality is strict for some of them.

Combining everything together, $\mathbb{E}_{\mathcal{M}_{CSM}} u_i(v_i, K_i) \geq \mathbb{E}_{\mathcal{M}_{CSM}} u_i(b_i, K_i)$ for all s_{-j} and the inequality is strict for some s_{-j}. Thus, fixing K_i in player i's strategy, it is dominant for i to report v_i and Claim 1 holds. □

It remains to show the following:

Claim 2. *Given that all players report their true values, it is dominant for a player i to report truthfully $K_i = (\mathcal{D}^i_j)_{j \neq i}$ as defined by the knowledge graph G: that is, $\mathcal{D}^i_j = \mathcal{D}_j$ for all j such that $(i, j) \in G$, and $\mathcal{D}^i_j = \perp$ otherwise.*

Proof. Arbitrarily fix a knowledge $K'_i = (\mathcal{D}'^i_j)_{j \neq i} \neq K_i$. Note that, under the no-bluff assumption, $\mathcal{D}'^i_j = \perp$ for all j such that $(i, j) \notin G$. Moreover, arbitrarily fix a strategy $s_j(v_j) = (v_j, K_j(v_j))$ for each $j \neq i$ and true value v_j. Note that j's reported knowledge depends on his true value and need not be truthful. We now compare $\mathbb{E}_{\mathcal{M}_{CSM}, \mathcal{D}_{-i}} u_i(v_i, K_i)$ and $\mathbb{E}_{\mathcal{M}_{CSM}, \mathcal{D}_{-i}} u_i(v_i, K'_i)$, where the expectation is taken over both the mechanism's random coins and the distributions of the other players' true values.

Same as before, conditional on $a = i$, the utility of i is equal to the reward he gets in Step 14. Since this step is reached by the mechanism with probability 1, given that it is reached, from i's point of view, v_{-i} is still distributed according to $\mathcal{D}_{-i}$, thus his expected reward is maximized by reporting K_i. That is,

$$\mathbb{E}_{v_{-i} \sim \mathcal{D}_{-i}}[u_i(v_i, K_i) \mid a = i] > \mathbb{E}_{v_{-i} \sim \mathcal{D}_{-i}}[u_i(v_i, K'_i) \mid a = i].$$

Next, we compare the two utilities conditional on $a \neq i$. Again, since Step 14 is reached with probability 1, the expected reward i gets in that step is strictly larger by reporting K_i than K'_i. Thus, it is enough to compare i's expected utilities at the end of Step 13, under (v_i, K_i) and (v_i, K'_i). Similar as before, $\mathcal{D}'_i$ is the same under both strategies, and the mechanism's execution is also the same till the round where $\mathcal{D}'_i$ is defined (or till the end if $\mathcal{D}'_i$ is not defined). There are three cases:

- First, if $\mathcal{D}'_i$ is not defined then neither K_i nor K'_i is used by the mechanism, thus i has the same utility under both strategies.

- Second, $\mathcal{D}'_i$ is defined and $i = i^*$ from that point on, then again K_i and K'_i are not used by the mechanism, thus i has the same utility under both strategies.

- Third, if $\mathcal{D}'_i$ is defined and $i \neq i^*$ in some round afterward, then i does not get the item under either strategy, thus his utility is 0 under both of them.

In sum, i's expected utility at the end of Step 13 is the same under both strategies. Summing it up with i's expected reward in Step 14, we have

$$\mathbb{E}_{v_{-i} \sim \mathcal{D}_{-i}}[u_i(v_i, K_i) \mid a \neq i] > \mathbb{E}_{v_{-i} \sim \mathcal{D}_{-i}}[u_i(v_i, K'_i) \mid a \neq i].$$

Combining the two inequalities above, we have

$$\mathbb{E}_{\mathcal{M}_{CSM}, \mathcal{D}_{-i}} u_i(v_i, K_i) > \mathbb{E}_{\mathcal{M}_{CSM}, \mathcal{D}_{-i}} u_i(v_i, K'_i),$$

thus Claim 2 holds. □

Lemma 1 follows directly from Claims 1 and 2. □

D Proofs for Section 4

D.1 Proof of Theorem 2

Theorem 2. (restated) *For any $k \in [n-1]$, any unit-demand auction instances $\hat{\mathcal{I}} = (N, M, \mathcal{D})$ and $\mathcal{I} = (N, M, \mathcal{D}, G)$ where G is k-bounded, the mechanism $\mathcal{M}_{CSUD}$ is 2-DST and*

$$\mathbb{E}_{v \sim \mathcal{D}} Rev(\mathcal{M}_{CSUD}(\mathcal{I})) \geq \frac{\tau_k}{6} \cdot OPT(\hat{\mathcal{I}}) - \epsilon.$$

To prove Theorem 2, we first prove the following lemma.

Lemma 2. *Mechanism $\mathcal{M}_{CSUD}$ is 2-DST.*

Proof. As before, the proof is done in two steps. First of all, arbitrarily fixing the players' reported knowledge, we show that it is dominant for the players to report their true valuations. Indeed, if player i is not in N_3 given the mechanism's randomness and the reported knowledge of all players, then his reported valuation is never used to compute his allocation or price, so he gets the same utility for reporting any valuation, including the true v_i.

If player i is in N_3, then his utility is determined by $\mathcal{M}_{UD}$. If $\mathcal{D}'_{ij}$ is defined to be 0 with probability 1 in Step 7, then i's reported value for item j is not given to $\mathcal{M}_{UD}$ as input, and i gets the same utility for reporting any value for j, including the true v_{ij}. Because i does not get such an item j, his utility is the same as an imaginary player $\hat{i}$ whose valuation is the same as i's, except that the true value of $\hat{i}$ for j is 0. Since $\mathcal{M}_{UD}$ is DST, no matter what the value distributions are, it is dominant for $\hat{i}$ to report his true valuation. Accordingly, it is dominant for i to report his true valuation v_i, concluding the first step of the proof.

Second of all, given that all players report their true valuations, we show that it is dominant for the players to report their true knowledge, as defined by G. Since player i's reported knowledge does not matter if he is in N_2, we focus on the case where he is in N_1. Notice that whether i's report is used to define $\mathcal{D}'_{i'j}$ for some $i' \in N_2$ and $j \in M$ solely depends on the players in N_1 who are before him. If $\mathcal{D}'_{i'j}$ has not been defined yet when i is processed, then i's expected reward is maximized by reporting $\mathcal{D}^i_{i'j} = \mathcal{D}_{i'j}$. Thus it is dominant for i to report his true knowledge, and Lemma 2 holds. □

Before analyzing the revenue of mechanism $\mathcal{M}_{CSUD}$, we first recall an important technique for reducing multi-parameter settings to single-parameter settings [16].

The COPIES setting. For a unit-demand auction instance $\hat{\mathcal{I}} = (N, M, \mathcal{D})$, the COPIES auction instance is defined as follows. We make m copies for each player, called *player copies*, and denote the resulting set by N^{CP}. We make n copies for each item, called *item copies*, and denote the resulting set by M^{CP}. Each player copy in N^{CP} is interested in a single item copy in M^{CP}, and two different player copies are interested in different item copies. More specifically, each player i's copy j has value v_{ij} for item j's copy i (and 0 for all other item copies), which is drawn independently from $\mathcal{D}_{ij}$.

The set of feasible allocations in the original unit-demand auction naturally defines the set of feasible allocations in the COPIES auction: for any feasible allocation A in the original setting, if player i gets item j, then in the corresponding allocation in the COPIES setting, player i's copy j gets item j's copy i. All other player copies get nothing. Since in the original setting each item is sold to at most one player, in the COPIES setting, for all copies of the same item, at most one of them is sold. Moreover, since each player gets one item in the original setting, in the COPIES setting, for all copies of the same player, at most one of them gets an item copy.

We denote by $\hat{\mathcal{I}}^{CP} = (N^{CP}, M^{CP}, \mathcal{D}^{CP})$ the COPIES instance. Intuitively, $\hat{\mathcal{I}}^{CP}$ involves more competition among the player copies: indeed, copies of the same original player are different players now, and their bids may be used to decide the prices for each other. Thus the optimal mechanism for COPIES might generate more revenue. The following is from [16] and restated using our conventions.

Lemma 3. [16] *For any deterministic DST Bayesian mechanism $\mathcal{M}$, $\mathbb{E}_{v\sim\mathcal{D}} Rev(\mathcal{M}(\hat{\mathcal{I}})) \le OPT(\hat{\mathcal{I}}^{CP})$.*

Now we are ready to analyze the revenue of mechanism $\mathcal{M}_{CSUD}$.

Proof of Theorem 2. Let $q = 1 - (k+1)^{-\frac{1}{k}}$. Following Lemma 2, it remains to show that

$$\mathbb{E}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{CSUD}(\mathcal{I})) \ge \frac{\tau_k}{6} \cdot OPT(\hat{\mathcal{I}}) - \epsilon.$$

Notice that for any player i and item j,

$$\begin{aligned} \Pr(\mathcal{D}_{ij} \text{ is reported in the mechanism}) &= \Pr(i \in N_2) \Pr(\exists i' \in N_1, (i,i') \in G_j \mid i \in N_2) \\ &\ge (1-q)(1-(1-q)^k), \end{aligned}$$

where the inequality is because G_j is k-bounded and the players are sampled independently. Taking derivatives of the last term, we have that it is maximized when $q = 1 - (k+1)^{-\frac{1}{k}}$, in which case

$$\Pr(\mathcal{D}_{ij} \text{ is reported in the mechanism}) \ge (k+1)^{-\frac{1}{k}} - (k+1)^{-\frac{k+1}{k}} = \tau_k.$$

Notice that τ_k is increasing in k, $\tau_1 = \frac{1}{4}$, and $\tau_k \to 1$ when k gets larger.

Below we use NM_3 to denote the set of player-item pairs whose distribution is reported in the mechanism: that is, the set of players N_3 together with their reported items. Accordingly,

$$\Pr((i,j) \in NM_3) \ge \tau_k. \tag{3}$$

Also, we use $\mathcal{D}_{NM_3} = \times_{(i,j)\in NM_3} \mathcal{D}_{ij}$ and $v_{NM_3} = (v_{ij})_{(i,j)\in NM_3}$ to denote the vector of distributions and the vector of true values for NM_3, respectively. Let $\hat{\mathcal{I}}_{NM_3} = (N_3, M, \mathcal{D}'_{N_3})$ be the unit-demand instance given to $\mathcal{M}_{UD}$ in Step 8. Note that $\mathcal{D}'_{NM_3} = \mathcal{D}_{NM_3}$ and $\mathcal{D}'_{ij}$ is constant 0 for any $(i,j) \notin NM_3$. Moreover, let R be the total reward given to the players using Brier's scoring rule in Step 7. It is easy to see that $R \le \epsilon$ always. Accordingly,

$$\begin{aligned} \mathop{\mathbb{E}}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{CSUD}(\mathcal{I})) &= \mathop{\mathbb{E}}_{NM_3} \mathop{\mathbb{E}}_{v_{NM_3}\sim\mathcal{D}_{NM_3}} Rev(\mathcal{M}_{UD}(\hat{\mathcal{I}}_{NM_3})) - R \\ &\ge \mathop{\mathbb{E}}_{NM_3} \mathop{\mathbb{E}}_{v_{NM_3}\sim\mathcal{D}_{NM_3}} Rev(\mathcal{M}_{UD}(\hat{\mathcal{I}}_{NM_3})) - \epsilon. \end{aligned}$$

Let $\hat{\mathcal{I}}^{CP}_{NM_3} = (N_3^{CP}, M^{CP}, \mathcal{D}'^{CP}_{N_3})$ be the COPIES setting corresponding to $\hat{\mathcal{I}}_{NM_3}$. By [35], there exists a Bayesian mechanism $\mathcal{M}^{CP}$ under the COPIES setting which achieves less revenue than $\mathcal{M}_{UD}$ under the unit-demand setting but is a 6-approximation to the optimal revenue in COPIES setting. That is, for any NM_3,

$$\mathop{\mathbb{E}}_{v_{NM_3}\sim\mathcal{D}_{NM_3}} Rev(\mathcal{M}_{UD}(\hat{\mathcal{I}}_{NM_3})) \geq \mathop{\mathbb{E}}_{v_{N_3^{CP}}\sim\mathcal{D}'^{CP}_{N_3}} Rev(\mathcal{M}^{CP}(\hat{\mathcal{I}}^{CP}_{NM_3})) \geq \frac{1}{6} \mathop{\mathbb{E}}_{v_{N_3^{CP}}\sim\mathcal{D}'^{CP}_{N_3}} OPT(\hat{\mathcal{I}}^{CP}_{NM_3}).$$

Combining the above two equations, we have

$$\mathop{\mathbb{E}}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{CSUD}(\mathcal{I})) \geq \frac{1}{6} \mathop{\mathbb{E}}_{NM_3} \mathop{\mathbb{E}}_{v_{N_3^{CP}}\sim\mathcal{D}'^{CP}_{N_3}} OPT(\hat{\mathcal{I}}^{CP}_{NM_3}) - \epsilon. \tag{4}$$

Below we focus on lower bounding $\mathop{\mathbb{E}}_{NM_3} \mathop{\mathbb{E}}_{v_{N_3^{CP}}\sim\mathcal{D}'^{CP}_{N_3}} OPT(\hat{\mathcal{I}}^{CP}_{NM_3})$. To do so, recall that $\hat{\mathcal{I}}^{CP} = (N^{CP}, M^{CP}, \mathcal{D}^{CP})$ is the COPIES setting corresponding to the original unit-demand auction $\hat{\mathcal{I}}$. For any set NM_3, let $OPT(\hat{\mathcal{I}}^{CP})_{NM_3}$ be the revenue of the optimal Bayesian mechanism on $\hat{\mathcal{I}}^{CP}$ projected to NM_3: that is, for any pair $(i,j) \notin NM_3$, the j-th copy of player i gets nothing and pays 0, even though he might get the i-th copy of item j according to $OPT(\hat{\mathcal{I}}^{CP})$. We first claim

$$\mathop{\mathbb{E}}_{NM_3} \mathop{\mathbb{E}}_{v_{N_3^{CP}}\sim\mathcal{D}'^{CP}_{N_3}} OPT(\hat{\mathcal{I}}^{CP}_{NM_3}) \geq \mathop{\mathbb{E}}_{NM_3} \mathop{\mathbb{E}}_{v_{N^{CP}}\sim\mathcal{D}^{CP}} OPT(\hat{\mathcal{I}}^{CP})_{NM_3}. \tag{5}$$

Indeed, for any NM_3, given the instance $\hat{\mathcal{I}}^{CP}_{NM_3}$, consider a Bayesian mechanism $\mathcal{M}'$ as follows:

- add m copies for each player $i \notin N_3$ and make each item into n copies;
- for any player $i \notin N_3$ and any item $j \in M$, let the value distribution of player i's copy j for item j's copy i be $\mathcal{D}_{ij}$;
- for any (i,j) such that $i \in N_3$ but $(i,j) \notin NM_3$, change the value distribution of i's copy j from $\mathcal{D}'_{ij}$ (which is constant 0) to $\mathcal{D}_{ij}$;
- for any $(i,j) \in NM_3$, keep the value distribution of i's copy j as $\mathcal{D}'_{ij} = \mathcal{D}_{ij}$;
- for any $(i,j) \in NM_3$, let i's copy j report his value, while for any $(i,j) \notin NM_3$, sample the value of i's copy j from $\mathcal{D}_{ij}$;
- run the optimal DST Bayesian mechanism on the resulting instance, which is exactly $\hat{\mathcal{I}}^{CP}$, with the reported and sampled values;
- project the resulting outcome to NM_3.

The key observation is that $\mathcal{M}'$ is a DST Bayesian mechanism for $\hat{\mathcal{I}}^{CP}_{NM_3}$. Indeed, for any (i,j) such that $i \in N_3$ and $(i,j) \notin NM_3$, the value of i's copy j in $\hat{\mathcal{I}}^{CP}_{NM_3}$ is constant 0, his reported value is not used by $\mathcal{M}'$, and at the end this player copy gets nothing and pays 0. Thus it is dominant for this player copy to report his true value 0. For any $(i,j) \in NM_3$, the utility of i's copy j under $\mathcal{M}'$ is the same as that under the optimal DST Bayesian mechanism. As it is dominant for this player copy to report his true value in the latter, it is so under $\mathcal{M}'$. By construction,

$$\mathop{\mathbb{E}}_{v_{N_3^{CP}}\sim\mathcal{D}'^{CP}_{N_3}} Rev(\mathcal{M}'(\hat{\mathcal{I}}^{CP}_{NM_3})) = \mathop{\mathbb{E}}_{v_{N^{CP}}\sim\mathcal{D}^{CP}} OPT(\hat{\mathcal{I}}^{CP})_{NM_3}.$$

Since

$$\mathop{\mathbb{E}}_{v_{N_3^{CP}} \sim \mathcal{D}'^{CP}_{N_3}} OPT(\hat{\mathcal{I}}^{CP}_{NM_3}) \geq \mathop{\mathbb{E}}_{v_{N_3^{CP}} \sim \mathcal{D}'^{CP}_{N_3}} Rev(\mathcal{M}'(\hat{\mathcal{I}}^{CP}_{NM_3})),$$

Equation 5 holds. Letting $P_{ij}(OPT(\hat{\mathcal{I}}^{CP}))$ be the price paid by player i's copy j under the optimal mechanism for $\hat{\mathcal{I}}^{CP}$, we can rewrite the right-hand side of Equation 5 as follows:

$$\begin{aligned}
&\mathop{\mathbb{E}}_{NM_3} \mathop{\mathbb{E}}_{v_{N^{CP}} \sim \mathcal{D}^{CP}} OPT(\hat{\mathcal{I}}^{CP})_{NM_3} = \mathop{\mathbb{E}}_{NM_3} \mathop{\mathbb{E}}_{v_{N^{CP}} \sim \mathcal{D}^{CP}} \sum_{(i,j) \in NM_3} P_{ij}(OPT(\hat{\mathcal{I}}^{CP})) \\
= &\mathop{\mathbb{E}}_{v_{N^{CP}} \sim \mathcal{D}^{CP}} \mathop{\mathbb{E}}_{NM_3} \sum_{(i,j) \in NM_3} P_{ij}(OPT(\hat{\mathcal{I}}^{CP})) \\
= &\mathop{\mathbb{E}}_{v_{N^{CP}} \sim \mathcal{D}^{CP}} \sum_{(i,j) \in N \times M} \Pr((i,j) \in NM_3) \cdot P_{ij}(OPT(\hat{\mathcal{I}}^{CP})), \qquad (6)
\end{aligned}$$

where the first equality is by the definition of revenue, the second is because the distribution of $v_{N^{CP}}$ does not depend on NM_3, and the third is by linearity of expectation and because $P_{ij}(OPT(\hat{\mathcal{I}}^{CP}))$ does not depend on NM_3. We can further lower bound the last term of Equation 6:

$$\begin{aligned}
&\mathop{\mathbb{E}}_{v_{N^{CP}} \sim \mathcal{D}^{CP}} \sum_{(i,j) \in N \times M} \Pr((i,j) \in NM_3) \cdot P_{ij}(OPT(\hat{\mathcal{I}}^{CP})) \\
\geq \ &\tau_k \mathop{\mathbb{E}}_{v_{N^{CP}} \sim \mathcal{D}^{CP}} \sum_{(i,j) \in N \times M} P_{ij}(OPT(\hat{\mathcal{I}}^{CP})) = \tau_k OPT(\hat{\mathcal{I}}^{CP}) \geq \tau_k OPT(\hat{\mathcal{I}}), \qquad (7)
\end{aligned}$$

where the first inequality is by Equation 3, the first equality is by the definition of revenue, and the second inequality is by Lemma 3.

Combining Equations 4, 5, 6 and 7, we have

$$\mathbb{E}_{v \sim \mathcal{D}} Rev(\mathcal{M}_{CSUD}(\mathcal{I})) \geq \frac{\tau_k}{6} \cdot OPT(\hat{\mathcal{I}}) - \epsilon,$$

and Theorem 2 holds. □

D.2 Proof of Theorem 3

Theorem 3. (restated) *For any $k \in [n-1]$, any additive auction instances $\hat{\mathcal{I}} = (N, M, \mathcal{D})$ and $\mathcal{I} = (N, M, \mathcal{D}, G)$ where G is k-bounded, the mechanism $\mathcal{M}_{CSA}$ is 2-DST and*

$$\mathbb{E}_{v \sim \mathcal{D}} Rev(\mathcal{M}_{CSA}(\mathcal{I})) \geq \max\{\frac{1}{24}, \frac{\tau_k}{12}\} \cdot OPT_B(\hat{\mathcal{I}}) - \epsilon.$$

To prove Theorem 3, we start by introducing mechanism $\mathcal{M}_{CSIM}$, which runs the following mechanism $\mathcal{M}_{CSIM,j}$ for each item j separately. Mechanism $\mathcal{M}_{CSIM,j}$ is similar to $\mathcal{M}_{CSUD}$, thus we have omitted many details.

We have the following lemma, whose proof is almost the same as that of Lemma 2 and is omitted.

Lemma 4. *Mechanism $\mathcal{M}_{CSIM,j}$ is 2-DST for each j, and mechanism $\mathcal{M}_{CSIM}$ is 2-DST as well.*

Next, we consider the expected revenue of $\mathcal{M}_{CSIM}$.

Mechanism 4. $\mathcal{M}_{CSIM,j}$

1: Each player i reports a value b_{ij} and a knowledge $K_{ij} = (\mathcal{D}^i_{i'j})_{i' \neq i}$.
2: Randomly partition the players into two sets, N_1 and N_2, where each player is independently put in N_1 with probability q and N_2 with probability $1-q$, with $q = 1-(k+1)^{-\frac{1}{k}}$.
3: Let N_3 be the set of players in N_2 whose distributions are reported by some players in N_1, and $\mathcal{D}'_{N_3,j}$ be the vector of reported distributions.
4: Reward players in N_1 using Brier's scoring rule, properly scaled so that the total reward R is at most ϵ/m.
5: Run Myerson's Mechanism on the single-good Bayesian instance $\hat{\mathcal{I}}_{N_3,j} = (N_3, \{j\}, \mathcal{D}'_{N_3,j})$ and the values being $(b_{ij})_{i \in N_3}$; and use the resulting allocation and prices to sell to players in N_3.

Lemma 5. $\mathbb{E}_{v \sim \mathcal{D}} Rev(\mathcal{M}_{CSIM}(\mathcal{I})) \geq \tau_k \mathbb{E}_{v \sim \mathcal{D}} Rev(IM(\hat{\mathcal{I}})) - \epsilon$.

Proof. Notice that

$$\mathbb{E}_{v \sim \mathcal{D}} Rev(\mathcal{M}_{CSIM}(\mathcal{I})) = \sum_{j \in M} \mathbb{E}_{(v_{ij})_{i \in N} \sim (\mathcal{D}_{ij})_{i \in N}} Rev(\mathcal{M}_{CSIM,j}(\mathcal{I}_j)),$$

where $\mathcal{I}_j = (N, \{j\}, (\mathcal{D}_{ij})_{i \in N}, G_j)$ is the single-good crowdsourced Bayesian instance with item j. Also notice that

$$\mathbb{E}_{v \sim \mathcal{D}} Rev(IM(\hat{\mathcal{I}})) = \sum_{j \in M} OPT(\hat{\mathcal{I}}_j),$$

where $\hat{\mathcal{I}}_j = (N, \{j\}, (\mathcal{D}_{ij})_{i \in N})$ is the single-good Bayesian instance with item j. Accordingly, letting $v_j = (v_{ij})_{i \in N}$ and $\mathcal{D}_j = (\mathcal{D}_{ij})_{i \in N}$ for each item j, it suffices to show that

$$\mathbb{E}_{v_j \sim \mathcal{D}_j} Rev(\mathcal{M}_{CSIM,j}(\mathcal{I}_j)) \geq \tau_k OPT(\hat{\mathcal{I}}_j) - \frac{\epsilon}{m}.$$

Using similar ideas and notations as in the proof of Theorem 2, we have that

$$\begin{aligned}
&\mathbb{E}_{v_j \sim \mathcal{D}_j} Rev(\mathcal{M}_{CSIM,j}(\mathcal{I}_j)) = \mathop{\mathbb{E}}_{N_3} \mathop{\mathbb{E}}_{v_{N_3,j} \sim \mathcal{D}_{N_3,j}} OPT(\hat{\mathcal{I}}_{N_3,j}) - R \\
\geq\ & \mathop{\mathbb{E}}_{N_3} \mathop{\mathbb{E}}_{v_{N_3,j} \sim \mathcal{D}_{N_3,j}} OPT(\hat{\mathcal{I}}_{N_3,j}) - \epsilon/m \geq \mathop{\mathbb{E}}_{N_3} \mathop{\mathbb{E}}_{v_j \sim \mathcal{D}_j} OPT(\hat{\mathcal{I}}_j)_{N_3} - \epsilon/m \\
=\ & \mathop{\mathbb{E}}_{N_3} \mathop{\mathbb{E}}_{v_j \sim \mathcal{D}_j} \sum_{i \in N_3} P_i(OPT(\hat{\mathcal{I}}_j)) - \epsilon/m = \mathop{\mathbb{E}}_{v_j \sim \mathcal{D}_j} \mathop{\mathbb{E}}_{N_3} \sum_{i \in N_3} P_i(OPT(\hat{\mathcal{I}}_j)) - \epsilon/m \\
=\ & \mathop{\mathbb{E}}_{v_j \sim \mathcal{D}_j} \sum_i \Pr(i \in N_3) \cdot P_i(OPT(\hat{\mathcal{I}}_j)) - \epsilon/m \geq \tau_k \mathop{\mathbb{E}}_{v_j \sim \mathcal{D}_j} \sum_i P_i(OPT(\hat{\mathcal{I}}_j)) - \epsilon/m \\
=\ & \tau_k OPT(\hat{\mathcal{I}}_j) - \epsilon/m,
\end{aligned}$$

as desired. Thus Lemma 5 holds. □

Notice that $\tau_k < \frac{1}{2}$ when $k \leq 3$. Below we show that a 2-approximation to mechanism IM can be achieved by using the 1-Lookahead mechanism [42] for single-good auctions, without random sampling. Mechanism $\mathcal{M}_{CS1LA}$ runs the following mechanism $\mathcal{M}_{CS1LA,j}$ for each item j separately.

Notice that $\mathcal{M}_{CS1LA,j}$ is not using the 1-Lookahead mechanism as a blackbox, because it has to handle boundary cases where the players' distributions are not all reported. As we will prove,

Mechanism 5. $\mathcal{M}_{CS1LA,j}$

1: Each player i reports a value b_{ij} and a knowledge $K_{ij} = (\mathcal{D}^i_{i'j})_{i' \neq i}$.
2: Reward each player by Brier's scoring rule, properly scaled so that the total reward R is at most ϵ/m.
3: Set $i^* = \arg\max_i b_{ij}$ and $p_{second} = \max_{i \neq i^*} b_{ij}$.
4: If i^*'s distribution is not reported, sell item j to him at price p_{second} and halt here.
5: Otherwise, let $\mathcal{D}'_{i^*j}$ be the reported distribution for i^* (if there are many reported distributions for him, then take the lexicographically first reporter).
6: Let $p_{i^*} = \max_p \Pr_{v_{i^*j} \sim \mathcal{D}'_{i^*j}} (v_{i^*j} \geq p \mid v_{i^*j} \geq p_{second}) \cdot p$. If $b_{i^*j} \geq p_{i^*}$ then sell item j to i^* at price p_{i^*}; otherwise the item is unsold.

the mechanism is 2-DST and all true distributions will indeed be reported by the players. However, for the mechanism to be well defined, it has to know what to do in all possible cases. Also notice that, running the 1-Lookahead mechanism on the set of players whose distributions are reported is not 2-DST. For example, if the player with the second highest value is the only one who knows the distribution for the player with the highest value, then he may choose not to report his knowledge about the latter, so that he himself has the highest value in the 1-Lookahead mechanism and gets a high utility. Still, the analysis of $\mathcal{M}_{CS1LA,j}$ strongly relies on the 1-Lookahead mechanism.

We have the following two lemmas.

Lemma 6. *Mechanism $\mathcal{M}_{CS1LA,j}$ is 2-DST for each j, and mechanism $\mathcal{M}_{CS1LA}$ is 2-DST as well.*

Proof. Arbitrarily fix an item j, a player i, a strategy subprofile of the other players and a knowledge of i. We show that $\mathcal{M}_{CS1LA,j}$ is *monotone* and uses the *threshold* payment for player i. Indeed, assume i gets the item by reporting a value b_{ij} and let him report a higher value b'_{ij}. Notice that b_{ij} is the highest value among all players, and so is b'_{ij}. If the other players did not report a distribution for i's value, then i's price is the second highest value, and he still gets the item at the same price by reporting b'_{ij}. Otherwise, i's price is the reserve price of the 1-Lookahead mechanism as defined in Step 6, which does not depend on his reported value, thus he still gets the item at this price by reporting b'_{ij}.

The above analysis also shows that player i pays the threshold payment, which is the second highest value if his distribution is not reported, while is the reserve price of the 1-Lookahead mechanism otherwise. Accordingly, it is dominant for i to report his true value.

Given that all players report their true values, it is easy to see that for each player i, it is dominant for i to report his true knowledge. Indeed, the only way for a player's reported knowledge to affect his own utility is to set his reward according to Brier's scoring rule, which is maximized by reporting his true knowledge.

Since the players have additive valuations and $\mathcal{M}_{CS1LA}$ runs each mechanism $\mathcal{M}_{CS1LA,j}$ separately, we have that $\mathcal{M}_{CS1LA}$ is 2-DST as well, and Lemma 6 holds. □

Lemma 7. $\mathbb{E}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{CS1LA}(\mathcal{I})) \geq \frac{1}{2}\mathbb{E}_{v\sim\mathcal{D}} Rev(IM(\hat{\mathcal{I}})) - \epsilon$.

Proof. Notice that, when the players report their true values and true knowledge, despite the reward given to the players, the outcome of each $\mathcal{M}_{CS1LA,j}$ on instance $\mathcal{I}_j = (N, \{j\}, \mathcal{D}_j, G_j)$ is exactly the same as that of the 1-Lookahead mechanism, denoted by mechanism $1LA$, running on instance $\hat{\mathcal{I}}_j = (N, \{j\}, \mathcal{D}_j)$. Accordingly,

$$\mathbb{E}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{CS1LA}(\mathcal{I})) = \sum_{j\in M} \mathbb{E}_{v_j\sim\mathcal{D}_j} Rev(\mathcal{M}_{CS1LA,j}(\mathcal{I}_j))$$
$$\geq \sum_{j\in M} \mathbb{E}_{v_j\sim\mathcal{D}_j} Rev(1LA(\hat{\mathcal{I}}_j)) - \epsilon \geq \sum_{j\in M} \frac{1}{2} OPT(\hat{\mathcal{I}}_j) - \epsilon = \frac{1}{2}\mathbb{E}_{v\sim\mathcal{D}} Rev(IM(\hat{\mathcal{I}})) - \epsilon,$$

where the second inequality is because the 1-Lookahead mechanism is a 2-approximation for the optimal Bayesian mechanism in single-good auctions [42]. Thus Lemma 7 holds. □

Remark. Notice that the approximation ratio of $\mathcal{M}_{CS1LA}$ does not depend on the specific value of k, as long as $k \geq 1$. Moreover, if the knowledge graphs are 2-connected, then no matter what k is, we can get a $(1 - \frac{1}{n})$-approximation to mechanism IM; see Corollary 1 at the end of Section 4.2.

Next, we describe the mechanism $\mathcal{M}_{CSBVCG}$, which approximates $BVCG$ in crowdsourced Bayesian settings.

Mechanism 6. $\mathcal{M}_{CSBVCG}$

1: Each player i reports a valuation $b_i = (b_{ij})_{j\in M}$ and a knowledge $K_i = (\mathcal{D}^i_{i'j})_{i'\neq i, j\in M}$.
2: Reward the players using Brier's scoring rule, such that the total reward R is at most ϵ.
3: For each item j, set $i^*(j) = \arg\max_i b_{ij}$ (ties broken lexicographically) and $p_j = \max_{i\neq i^*} b_{ij}$.
4: **for** each player i **do**
5: Let $M_i = \{j \mid i^*(j) = i\}$.
6: If not all m distributions of i's values are reported, i gets no item and items in M_i are unsold.
7: Otherwise, let $\mathcal{D}'_i$ be the vector of reported distributions for i's values and compute the entry fee $e_i(\mathcal{D}'_i, b_{-i})$ using $BVCG$.
8: If $\sum_{j\in M_i} b_{ij} \geq e_i(\mathcal{D}'_i, b_{-i}) + \sum_{j\in M_i} p_j$ then i gets M_i with price $e_i(\mathcal{D}'_i, b_{-i}) + \sum_{j\in M_i} p_j$; otherwise i gets no item and items in M_i are unsold.
9: **end for**

We have the following two lemmas.

Lemma 8. *Mechanism $\mathcal{M}_{CSBVCG}$ is 2-DST.*

Proof. Arbitrarily fix a player i, a strategy subprofile of the other players, and a knowledge of i. Notice that if not all m distributions of i's values are reported by others, then i gets nothing and pays nothing, so it does not matter what valuation he reports about himself. Otherwise, $\mathcal{M}_{CSBVCG}$ sells to player i in the same way as how $BVCG$ would sell to him: using the other players' highest reported values as reserve price for each item, either i gets the whole set of items for which his value passes the reserve price, or he gets nothing and those items are unsold to anybody. Following [47], it is dominant for i to report his true values given any entry fee that does not depend on his reported values, so truth-telling is still dominant for i when the entry fee is computed based on $\mathcal{D}'_i$ and the reserve prices.

Furthermore, notice that a player i's reported knowledge K_i about others affects neither M_i nor e_i. Thus, K_i is only used to compute player i's reward based on Brier's scoring rule, and i's expected reward is maximized by reporting his true knowledge, given that all players report their true values. □

Lemma 9. $\mathbb{E}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{CSBVCG}(\mathcal{I})) \geq \mathbb{E}_{v\sim\mathcal{D}} Rev(BVCG(\hat{\mathcal{I}})) - \epsilon.$

Proof. Since $\mathcal{M}_{CSBVCG}$ retrieves the whole distribution $\mathcal{D}$ from the players, its outcome is exactly the same as that of $BVCG$ under the Bayesian instance $\hat{\mathcal{I}}$. Thus, its expected revenue is that of $BVCG$ minus the total reward, which is at most ϵ. □

Remark. Similar to $\mathcal{M}_{CS1LA}$, the revenue guarantee of $\mathcal{M}_{CSBVCG}$ does not depend on k. Indeed, notice the special structures of the two Bayesian mechanisms $1LA$ and $BVCG$: the potential set of items a player may get solely depends on the players' values, and the distributions are only used to compute better reserve prices to increase revenue. Therefore, in the crowdsourced setting we can allow a player to be both a reporter about the others' distributions and a potential winner of some items. In some other mechanisms, e.g., Myerson's mechanism, the distributions are used both to choose the potential winner and to set his price, thus in the crowdsourced setting we must separate the knowledge reporters and the potential winners. Indeed, if player i would have gotten some items when he reports nothing about player j, but those items will be given to j if he reports truthfully, then i may be better off reporting nothing. That is why random sampling is needed in these cases.

Finally, our mechanism $\mathcal{M}_{CSA}$ is defined as follows. It flips a fair coin; if heads comes up then it runs $\mathcal{M}_{CSBVCG}$; and if tails comes up, then it runs $\mathcal{M}_{CS1LA}$ when $k \leq 3$ and $\mathcal{M}_{CSIM}$ when $k > 3$.

Proof of Theorem 3. The mechanism $\mathcal{M}_{CSA}$ is clearly 2-DST, since all the sub-mechanisms are 2-DST and which mechanism is chosen does not depend on the players' strategies.

When $k \leq 3$, $\tau_k < \frac{1}{2}$, and by Lemmas 7 and 9 we have

$$\begin{aligned}\mathop{\mathbb{E}}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{CSA}(\mathcal{I})) &= \frac{1}{2}\mathop{\mathbb{E}}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{CSBVCG}(\mathcal{I})) + \frac{1}{2}\mathop{\mathbb{E}}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{CS1LA}(\mathcal{I})) \\ &\geq \frac{1}{2}\mathop{\mathbb{E}}_{v\sim\mathcal{D}} Rev(BVCG(\hat{\mathcal{I}})) + \frac{1}{4}\mathop{\mathbb{E}}_{v\sim\mathcal{D}} Rev(IM(\hat{\mathcal{I}})) - \epsilon.\end{aligned}$$

When $k > 3$, $\tau_k \geq \frac{1}{2}$, and by Lemmas 5 and 9 we have

$$\begin{aligned}\mathop{\mathbb{E}}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{CSA}(\mathcal{I})) &= \frac{1}{2}\mathop{\mathbb{E}}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{CSBVCG}(\mathcal{I})) + \frac{1}{2}\mathop{\mathbb{E}}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{CSIM}(\mathcal{I})) \\ &\geq \frac{1}{2}\mathop{\mathbb{E}}_{v\sim\mathcal{D}} Rev(BVCG(\hat{\mathcal{I}})) + \frac{\tau_k}{2}\mathop{\mathbb{E}}_{v\sim\mathcal{D}} Rev(IM(\hat{\mathcal{I}})) - \epsilon.\end{aligned}$$

Thus for all k, $\mathop{\mathbb{E}}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{CSA}(\mathcal{I})) \geq \frac{1}{2}\mathop{\mathbb{E}}_{v\sim\mathcal{D}} Rev(BVCG(\hat{\mathcal{I}})) + \max\left\{\frac{1}{4}, \frac{\tau_k}{2}\right\}\mathop{\mathbb{E}}_{v\sim\mathcal{D}} Rev(IM(\hat{\mathcal{I}})) - \epsilon$.

Following [13], $OPT_B(\hat{\mathcal{I}}) \leq 2\mathop{\mathbb{E}}_{v\sim\mathcal{D}} Rev(BVCG(\hat{\mathcal{I}})) + 6\mathop{\mathbb{E}}_{v\sim\mathcal{D}} Rev(IM(\hat{\mathcal{I}}))$. Therefore

$$\mathop{\mathbb{E}}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{CSA}(\mathcal{I})) \geq \max\left\{\frac{1}{24}, \frac{\tau_k}{12}\right\} \cdot OPT_B(\hat{\mathcal{I}}) - \epsilon,$$

and Theorem 3 holds. □

E Proofs for Section 5

Theorem 4. (restated) *For any $k \in [n-1]$, for any auction instances $\hat{\mathcal{I}} = (N, M, \mathcal{D})$ and $\mathcal{I} = (N, M, \mathcal{D}, G)$ where G is k-bounded, if $\mathcal{M}_B$ is DST then $\mathcal{M}_{CSB}$ is 2-DST; and if $\mathcal{M}_B$ is BIC then $\mathcal{M}_{CSB}$ is BIC. Moreover, if $\mathcal{M}_B$ is a σ-approximation to OPT (respectively, OPT_B), then $\mathcal{M}_{CSB}$ is a $\tau_k\sigma$-approximation to OPT (respectively, OPT_B), up to an additive ϵ discount.*

Proof. The fact that $\mathcal{M}_B$ being DST implies $\mathcal{M}_{CSB}$ being 2-DST is proved in the same way as previous mechanisms such as $\mathcal{M}_{CSUD}$, thus all the details are omitted.

If $\mathcal{M}_B$ is BIC, then we need to show that $\mathcal{M}_{CSB}$ is also BIC: that is, all players reporting their true valuations and true knowledge is a Bayesian Nash equilibrium. This is almost immediate by definition. Indeed, given that all players but player i tell the truth about their valuations and knowledge, if player i is in N_1, then his reported valuation is not used and his reported knowledge only affects his utility via Brier's scoring rule, thus telling the truth about both maximizes i's expected utility. If player i is in N_2, then his reported knowledge is never used and surely telling the truth is utility-maximizing. Moreover, given $i \in N_2$, whether $i \in N_3$ or not solely depends on the other players' strategies. If $i \notin N_3$, then i's utility is 0 no matter what valuation he reports. If $i \in N_3$ then, given that $\mathcal{M}_B$ is BIC, $\mathcal{D}'_{N_3} = \mathcal{D}_{N_3}$ and all the others report their true valuations, reporting his own true valuation maximizes i's expected utility. Therefore $\mathcal{M}_{CSB}$ is BIC.

Below we analyze the revenue of $\mathcal{M}_{CSB}$ when $\mathcal{M}_B$ is a σ-approximation to OPT. The case for OPT_B is exactly the same and thus omitted. More precisely, we show

$$\mathbb{E}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{CSB}(\mathcal{I})) \geq \tau_k \sigma OPT(\hat{\mathcal{I}}) - \epsilon,$$

using ideas and notations similar to those in the proof of Theorem 2.

First, for each player i,

$$\Pr(i \in N_3) = \Pr(i \in N_2)\Pr(\exists i' \in N_1 \text{ s.t. } (i', i) \in G \mid i \in N_2) \geq (1-q)(1-(1-q)^k) = \tau_k.$$

Accordingly,

$$\begin{aligned}
&\mathop{\mathbb{E}}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{CSB}(\mathcal{I})) = \mathop{\mathbb{E}}_{N_3} \mathop{\mathbb{E}}_{v_{N_3}\sim\mathcal{D}_{N_3}} Rev(\mathcal{M}_B(\hat{\mathcal{I}}_{N_3})) - R \geq \mathop{\mathbb{E}}_{N_3} \sigma OPT(\hat{\mathcal{I}}_{N_3}) - \epsilon \\
\geq\ & \sigma \mathop{\mathbb{E}}_{N_3} OPT(\hat{\mathcal{I}})_{N_3} - \epsilon = \sigma \mathop{\mathbb{E}}_{N_3} \mathop{\mathbb{E}}_{v\sim\mathcal{D}} \sum_{i\in N_3} P_i(OPT(\hat{\mathcal{I}})) - \epsilon = \sigma \mathop{\mathbb{E}}_{v\sim\mathcal{D}} \mathop{\mathbb{E}}_{N_3} \sum_{i\in N_3} P_i(OPT(\hat{\mathcal{I}})) - \epsilon \\
=\ & \sigma \mathop{\mathbb{E}}_{v\sim\mathcal{D}} \sum_{i\in N} \Pr(i\in N_3) P_i(OPT(\hat{\mathcal{I}})) - \epsilon \geq \tau_k \sigma \mathop{\mathbb{E}}_{v\sim\mathcal{D}} \sum_{i\in N} P_i(OPT(\hat{\mathcal{I}})) - \epsilon = \tau_k \sigma OPT(\hat{\mathcal{I}}) - \epsilon,
\end{aligned}$$

as desired. Thus Theorem 4 holds. □

F When Not Everything is Known

Recall that when not everything is known, the revenue benchmarks for crowdsourced Bayesian mechanisms should take into consideration the missing knowledge. More precisely, letting $\mathcal{I} = (N, \mathcal{D}, G)$ be a single-good crowdsourced auction instance and $N' = \{i \mid \exists i' \text{ s.t. } (i', i) \in G\}$, the benchmark is $OPT(N', \mathcal{D}_{N'})$. For unit-demand auction and additive auction instances $\mathcal{I} = (N, M, \mathcal{D}, G)$, let $\mathcal{D}' = \times_{i\in N, j\in M} \mathcal{D}'_{ij}$ where $\mathcal{D}'_{ij} = \mathcal{D}_{ij}$ if there exists player i' such that $(i', i) \in G_j$, and $\mathcal{D}'_{ij}$ is constant 0 otherwise. We say that $\mathcal{D}'$ is $\mathcal{D}$ *projected on* G. Then the benchmark is $OPT(N, M, \mathcal{D}')$. The benchmarks for crowdsourced auctions with player-wise information are defined similarly.

Under the no-bluff assumption, for all random sampling mechanisms we have constructed before, when not every distribution is known by some one, we can simply set $q = \frac{1}{2}$ and keep all the rest unchanged. Through exactly the same analysis as before, we have the following two theorems, with proofs omitted.

Theorem 6. *In the partial information setting, for any unit-demand auction instances* $\hat{\mathcal{I}} = (N, M, \mathcal{D})$, $\mathcal{I} = (N, M, \mathcal{D}, G)$ *and* $\mathcal{I}' = (N, M, \mathcal{D}')$ *where* $\mathcal{D}'$ *is* $\mathcal{D}$ *projected on* G*, There exists a crowdsourced mechanism* $\mathcal{M}'_{CSUD}$ *that is 2-DST and* $\mathbb{E}_{v\sim\mathcal{D}} Rev(\mathcal{M}'_{CSUD}(\mathcal{I})) \geq \frac{1}{24} \cdot OPT(\mathcal{I}') - \epsilon$.

Theorem 7. *In the player-wise information setting, for any auction instances $\hat{\mathcal{I}} = (N, M, \mathcal{D})$, $\mathcal{I} = (N, M, \mathcal{D}, G)$ and $\mathcal{I}' = (N, M, \mathcal{D}')$ where $\mathcal{D}'$ is $\mathcal{D}$ projected on G, for any Bayesian mechanism $\mathcal{M}_B$, there exists a mechanism $\mathcal{M}'_{CSB}$ such that: if $\mathcal{M}_B$ is DST then $\mathcal{M}'_{CSB}$ is 2-DST; and if $\mathcal{M}_B$ is BIC then $\mathcal{M}'_{CSB}$ is BIC. Moreover, if $\mathcal{M}_B$ is a σ-approximation to OPT (respectively, OPT_B), then $\mathcal{M}'_{CSB}(\mathcal{I})$ is a $\frac{\sigma}{4}$-approximation to $OPT(\mathcal{I}')$ (respectively, $OPT_B(\mathcal{I}')$), up to an additive ϵ discount.*

The case of additive auctions with partial information is more complex. To begin with, the crowdsourced Individual Myerson mechanism $\mathcal{M}_{CSIM}$ can be changed in a similar way, and the 1-Lookahead mechanism $\mathcal{M}_{CS1LA}$ does not need any change. Given their corresponding approximation ratios, actually we will only need $\mathcal{M}_{CS1LA}$.

The mechanism $\mathcal{M}_{CSBVCG}$ requires extra effort to ensure that it still achieves a constant approximation to $BVCG$ on the projected instance. We define the altered mechanism $\mathcal{M}'_{CSBVCG}$ in Mechanism 7.

Mechanism 7. $\mathcal{M}'_{CSBVCG}$

Each player i reports a valuation $b_i = (b_{ij})_{j \in M}$ and a knowledge $K_i = (\mathcal{D}^i_{i'j})_{i' \neq i, j \in M}$.
Reward the players using Brier's scoring rule, such that the total reward R is at most ϵ.
For each item j, set $i^*(j) = \arg\max_i b_{ij}$ (ties broken lexicographically) and $p_j = \max_{i \neq i^*} b_{ij}$.
for each player i **do**
    Let $M_i = \{j \mid i^*(j) = i\}$.
    Partition M into M_i^1 and M_i^2 as follows: $\forall j \in M_i^1$, some i' has reported $\mathcal{D}^{i'}_{ij} \neq \bot$ (if there are more than one reporters, take the lexicographically first); and $\forall j \in M_i^2$, $\mathcal{D}^{i'}_{ij} = \bot$ for all i'.
    $\forall j \in M_i^1$, set $\mathcal{D}'_{ij} = \mathcal{D}^{i'}_{ij}$; and $\forall j \in M_i^2$, set $\mathcal{D}'_{ij}$ to be 0 with probability 1.
    Compute the entry fee $e_i(\mathcal{D}'_i, b_{-i})$ using $BVCG$.
    If $\sum_{j \in M_i^1 \cap M_i} b_{ij} \geq e_i + \sum_{j \in M_i^1 \cap M_i} p_j$, then i gets $M_i^1 \cap M_i$ with price $e_i + \sum_{j \in M_i^1 \cap M_i} p_j$; otherwise the items in $M_i^1 \cap M_i$ are not sold to anybody.
    In addition, sell each item j in $M_i^2 \cap M_i$ to player i with price p_j.
end for

The mechanism $\mathcal{M}'_{CSA}$ randomly chooses between $\mathcal{M}'_{CSBVCG}$ and $\mathcal{M}_{CS1LA}$ with equal probability. We have the following theorem.

Theorem 8. *For any additive auction instances $\hat{\mathcal{I}} = (N, M, \mathcal{D})$, $\mathcal{I} = (N, M, \mathcal{D}, G)$ and $\mathcal{I}' = (N, M, \mathcal{D}')$ where $\mathcal{D}'$ is $\mathcal{D}$ projected on G, $\mathcal{M}'_{CSA}$ is 2-DST and $\mathbb{E}_{v \sim \mathcal{D}} Rev(\mathcal{M}'_{CSA}(\mathcal{I})) \geq \frac{1}{24} \cdot OPT_B(\mathcal{I}') - \epsilon$.*

Proof. We focus on analyzing $\mathcal{M}'_{CSBVCG}$, as the other part of the analysis is almost the same as Theorem 3. The main difference between the two mechanisms is that, when a distribution about player i's value for item j is not report, $\mathcal{M}_{CSBVCG}$ chooses not to sell anything to player i and charges him with price 0, while $\mathcal{M}'_{CSBVCG}$ divides the item set into two subsets M_i^1 and M_i^2 and sells them using the optimal entry fee and the second-price mechanism, respectively.

It is easy to verify that mechanism $\mathcal{M}'_{CSBVCG}$ is 2-DST. Indeed, given the distributions and values reported by other players, M_i^1 and M_2^2 do not depend on i's strategy at all, neither does the entry fee for M_i^1. Thus, it is dominate for i to report his true values for M_i^1, following $BVCG$; and it is dominant for him to report his true values for M_i^2, following the second-price mechanism. Given that all players truthfully report their values, it is dominate for each i to report his true knowledge. Indeed, the set M_i only depends on the reported values. i's reported knowledge does not affect whether he gets some items or not, and only affects his utility via Brier's scoring rule.

Next, we lower bound the expected revenue of $\mathcal{M}'_{CSBVCG}$. We start with the case where G is such that only a single distribution $\mathcal{D}_{ij}$ is unknown to all players, and we will generalize to arbitrary number of unknown distributions afterward. We have the following.

Lemma 10. $\mathbb{E}_{v\sim\mathcal{D}}Rev(\mathcal{M}'_{CSBVCG}(\mathcal{I})) \geq \frac{1}{2}\mathbb{E}_{v'\sim\mathcal{D}'}Rev(BVCG(\mathcal{I}')) - \epsilon$.

Proof. Notice that the $-\epsilon$ part comes from the total reward given by the scoring rule in $\mathcal{M}'_{CSBVCG}$, and we will ignore it below, treating $Rev(\mathcal{M}'_{CSBVCG}(\mathcal{I}))$ as the revenue without giving out the rewards. Since

$$\mathbb{E}_{v\sim\mathcal{D}}Rev(\mathcal{M}'_{CSBVCG}(\mathcal{I})) = \sum_{i'\in N}\mathbb{E}_{v\sim\mathcal{D}}P_{i'}(\mathcal{M}'_{CSBVCG}(\mathcal{I}))$$

and

$$\mathbb{E}_{v'\sim\mathcal{D}'}Rev(BVCG(\mathcal{I}')) = \sum_{i'\in N}\mathbb{E}_{v'\sim\mathcal{D}'}P_{i'}(BVCG(\mathcal{I}')),$$

we compare the corresponding terms in the two summations player by player.

First, for player i, given any valuation subprofile v_{-i}, both mechanisms compute the optimal entry fee for i based on $(\mathcal{D}_{ij'})_{j'\neq j}$ and $(v_{i'j'})_{i'\neq i,j'\neq j}$. Thus they have the same reserve prices and entry fee for i, and generate the same expected revenue from i on items $-j$, which is all the revenue $BVCG(\mathcal{I}')$ generates from i as it never sells j to i. Differently, $\mathcal{M}'_{CSBVCG}(\mathcal{I})$ sells j to i by the second-price mechanism, and generates extra revenue whenever $i = \arg\max_{\hat{i}} v_{\hat{i}j}$ (ties broken lexicographically). Therefore

$$\begin{aligned}
&\mathbb{E}_{v\sim\mathcal{D}}P_i(\mathcal{M}'_{CSBVCG}(\mathcal{I})) = \mathbb{E}_{v'\sim\mathcal{D}'}P_i(BVCG(\mathcal{I}')) + \sum_{v:\ i=\arg\max_{\hat{i}} v_{\hat{i}j}} \Pr_{\mathcal{D}}(v)\cdot\max_{\hat{i}\neq i} v_{\hat{i}j}\\
=\ &\mathbb{E}_{v'\sim\mathcal{D}'}P_i(BVCG(\mathcal{I}')) + \sum_{i'\neq i}\ \sum_{v:i=\arg\max_{\hat{i}} v_{\hat{i}j},i'=\arg\max_{\hat{i}\neq i} v_{\hat{i}j}} \Pr_{\mathcal{D}}(v)\cdot v_{i'j}. \qquad (8)
\end{aligned}$$

Now arbitrarily fix a player $i' \neq i$ and let

$$R_{i'} = \sum_{v:i=\arg\max_{\hat{i}} v_{\hat{i}j},i'=\arg\max_{\hat{i}\neq i} v_{\hat{i}j}} \Pr_{\mathcal{D}}(v)\cdot v_{i'j}.$$

We make the following claim.

Claim 3. $\mathbb{E}_{v\sim\mathcal{D}}P_{i'}(\mathcal{M}'_{CSBVCG}(\mathcal{I})) \geq \frac{1}{2}\left(\mathbb{E}_{v'\sim\mathcal{D}'}P_{i'}(BVCG(\mathcal{I}')) - R_{i'}\right)$.

Leaving the proof of Claim 3 for a moment, first notice that, combining Equation 8 and Claim 3, we have

$$\begin{aligned}
&\mathbb{E}_{v\sim\mathcal{D}}Rev(\mathcal{M}'_{CSBVCG}(\mathcal{I})) = \mathbb{E}_{v\sim\mathcal{D}}P_i(\mathcal{M}'_{CSBVCG}(\mathcal{I})) + \sum_{i'\neq i}\mathbb{E}_{v\sim\mathcal{D}}P_{i'}(\mathcal{M}'_{CSBVCG}(\mathcal{I}))\\
\geq\ &\frac{1}{2}\left(\mathbb{E}_{v'\sim\mathcal{D}'}P_i(BVCG(\mathcal{I}')) + \sum_{i'\neq i}R_{i'}\right) + \sum_{i'\neq i}\frac{1}{2}\left(\mathbb{E}_{v'\sim\mathcal{D}'}P_{i'}(BVCG(\mathcal{I}')) - R_{i'}\right)\\
=\ &\frac{1}{2}\sum_{i'\in N}\mathbb{E}_{v'\sim\mathcal{D}'}P_{i'}(BVCG(\mathcal{I}')) = \frac{1}{2}\mathbb{E}_{v'\sim\mathcal{D}'}Rev(BVCG(\mathcal{I}')),
\end{aligned}$$

and Lemma 10 holds. □

It remains to prove Claim 3.

Proof of Claim 3. To do so, arbitrarily fix a valuation subprofile $v_{-i'}$ and let $v'_{-i'}$ be the same as $v_{-i'}$ except $v'_{ij} = 0$. Since

$$\mathbb{E}_{v\sim\mathcal{D}} P_{i'}(\mathcal{M}'_{CSBVCG}(\mathcal{I})) = \sum_{v_{-i'}} \Pr_{\mathcal{D}_{-i'}}(v_{-i'})\ \mathbb{E}_{v_{i'}\sim\mathcal{D}_{i'}} P_{i'}(\mathcal{M}'_{CSBVCG}(\mathcal{I}) \mid v_{-i'})$$

and

$$\mathbb{E}_{v'\sim\mathcal{D}'} P_{i'}(BVCG(\mathcal{I}')) = \sum_{v_{-i'}} \Pr_{\mathcal{D}_{-i'}}(v_{-i'})\ \mathbb{E}_{v_{i'}\sim\mathcal{D}_{i'}} P_{i'}(BVCG(\mathcal{I}') \mid v'_{-i'}),$$

We compare the internal expectations, $\mathbb{E}_{v_{i'}\sim\mathcal{D}_{i'}} P_{i'}(\mathcal{M}'_{CSBVCG}(\mathcal{I}) \mid v_{-i'})$ and $\mathbb{E}_{v_{i'}\sim\mathcal{D}_{i'}} P_{i'}(BVCG(\mathcal{I}') \mid v'_{-i'})$, for each $v_{-i'}$ and the corresponding $v'_{-i'}$. Let $\overline{v_{-i'}}$ be the vector of reserve prices for i', that is, $\overline{v_{-i'}}_k = \max_{\hat{i}\neq i'} v_{\hat{i}k}$ for each item $k \in M$; and let $\overline{v'_{-i'}}$ be defined similarly. We distinguish two cases.

First, if $v_{ij} \leq \max_{\hat{i}\neq i,i'} v_{\hat{i}j}$, then $\overline{v_{-i'}} = \overline{v'_{-i'}}$, $\mathcal{M}'_{CSBVCG}(\mathcal{I})$ and $BVCG(\mathcal{I}')$ have the same reserve prices and entry fee for i', so

$$\mathbb{E}_{v_{i'}\sim\mathcal{D}_{i'}} P_{i'}(\mathcal{M}'_{CSBVCG}(\mathcal{I}) \mid v_{-i'}) = \mathbb{E}_{v_{i'}\sim\mathcal{D}_{i'}} P_{i'}(BVCG(\mathcal{I}') \mid v'_{-i'}). \tag{9}$$

Second, if $v_{ij} > \max_{\hat{i}\neq i,i'} v_{\hat{i}j}$, then $\overline{v_{-i'}}$ and $\overline{v'_{-i'}}$ differ at the j-th component: $\overline{v_{-i'}}_j = v_{ij}$ while $\overline{v'_{-i'}}_j = \max_{\hat{i}\neq i,i'} v_{\hat{i}j}$. We write $\mathbb{E}_{v_{i'}\sim\mathcal{D}_{i'}} P_{i'}(BVCG(\mathcal{I}') \mid v'_{-i'})$ as the summation of three parts:

$$\begin{aligned}
&\mathbb{E}_{v_{i'}\sim\mathcal{D}_{i'}} P_{i'}(BVCG(\mathcal{I}') \mid v'_{-i'}) \\
=\ & \sum_{v_{i'j}:\ i'\neq\arg\max_{\hat{i}\neq i} v_{\hat{i}j}} \Pr_{\mathcal{D}_{i'j}}(v_{i'j}) \cdot \mathop{\mathbb{E}}_{(v_{i'j'})_{j'\neq j}\sim(\mathcal{D}_{i'j'})_{j'\neq j}} P_{i'}(BVCG(\mathcal{I}') \mid v'_{-i'}, v_{i'j}) && (10)\\
+\ & \sum_{\substack{v_{i'j}:\ i'=\arg\max_{\hat{i}\neq i} v_{\hat{i}j},\\ i'\neq\arg\max_{\hat{i}\in N} v_{\hat{i}j}}} \Pr_{\mathcal{D}_{i'j}}(v_{i'j}) \cdot \mathop{\mathbb{E}}_{(v_{i'j'})_{j'\neq j}\sim(\mathcal{D}_{i'j'})_{j'\neq j}} P_{i'}(BVCG(\mathcal{I}') \mid v'_{-i'}, v_{i'j}) && (11)\\
+\ & \sum_{v_{i'j}:\ i'=\arg\max_{\hat{i}\in N} v_{\hat{i}j}} \Pr_{\mathcal{D}_{i'j}}(v_{i'j}) \cdot \mathop{\mathbb{E}}_{(v_{i'j'})_{j'\neq j}\sim(\mathcal{D}_{i'j'})_{j'\neq j}} P_{i'}(BVCG(\mathcal{I}') \mid v'_{-i'}, v_{i'j}). && (12)
\end{aligned}$$

Below we upper bound each part by a proper function of $\mathbb{E}_{v_{i'}\sim\mathcal{D}_{i'}} P_{i'}(\mathcal{M}'_{CSBVCG}(\mathcal{I}) \mid v_{-i'})$.

For each $v_{i'j}$ such that $i' \neq \arg\max_{\hat{i}\neq i} v_{\hat{i}j}$, that is, those in Equation 10, i' is not the potential winner for j under $BVCG(\mathcal{I}')$. Let $e^B_{i'}(\mathcal{D}_{i'}, v'_{-i'})$ be the entry fee of i' under $BVCG(\mathcal{I}')$. We have that $\mathop{\mathbb{E}}_{(v_{i'j'})_{j'\neq j}\sim(\mathcal{D}_{i'j'})_{j'\neq j}} P_{i'}(BVCG(\mathcal{I}') \mid v'_{-i'}, v_{i'j})$ is the same as the expected revenue generated by selling items in $N \setminus \{j\}$ to i' under reserve prices $\overline{v'_{-i'}}_{-j} = \overline{v_{-i'}}_{-j}$ and entry fee $e^B_{i'}(\mathcal{D}_{i'}, v'_{-i'})$. That is, denoting the latter as mechanism $M_{i',\ N\setminus\{j\},\ \overline{v_{-i'}}_{-j},\ e^B_{i'}(\mathcal{D}_{i'},v'_{-i'})}$, we have

$$\mathop{\mathbb{E}}_{(v_{i'j'})_{j'\neq j}\sim(\mathcal{D}_{i'j'})_{j'\neq j}} P_{i'}(BVCG(\mathcal{I}') \mid v'_{-i'}, v_{i'j}) = \mathop{\mathbb{E}}_{(v_{i'j'})_{j'\neq j}\sim(\mathcal{D}_{i'j'})_{j'\neq j}} P_{i'}(M_{i',\ N\setminus\{j\},\ \overline{v_{-i'}}_{-j},\ e^B_{i'}(\mathcal{D}_{i'},v'_{-i'})}).$$

It is easy to see that

$$\mathop{\mathbb{E}}_{(v_{i'j'})_{j'\neq j}\sim(\mathcal{D}_{i'j'})_{j'\neq j}} P_{i'}(M_{i',\ N\setminus\{j\},\ \overline{v_{-i'}}_{-j},\ e^B_{i'}(\mathcal{D}_{i'},v'_{-i'})}) \leq \mathop{\mathbb{E}}_{v_{i'}\sim\mathcal{D}_{i'}} P_{i'}(M_{i',\ N,\ \overline{v_{-i'}},\ e^B_{i'}(\mathcal{D}_{i'},v'_{-i'})}), \tag{13}$$

since whenever the first mechanism makes a sale to i', the second mechanism makes a sale to i' with at least the same revenue, and possibly more from item j. Moreover,

$$\mathop{\mathbb{E}}_{v_{i'}\sim\mathcal{D}_{i'}} P_{i'}(M_{i',\ N,\ \overline{v_{-i'}},\ e^B_{i'}(\mathcal{D}_{i'},v'_{-i'})}) \leq \mathbb{E}_{v_{i'}\sim\mathcal{D}_{i'}} P_{i'}(\mathcal{M}'_{CSBVCG}(\mathcal{I}) \mid v_{-i'}), \tag{14}$$

because $\mathcal{M}'_{CSBVCG}$ computes the optimal entry fee under the same reserve prices so as to maximize the expected revenue from j. Combining the above three equations together, we have

$$\mathop{\mathbb{E}}_{(v_{i'j'})_{j'\neq j}\sim(\mathcal{D}_{i'j'})_{j'\neq j}} P_{i'}(BVCG(\mathcal{I}') \mid v'_{-i'}, v_{i'j}) \leq \mathbb{E}_{v_{i'}\sim\mathcal{D}_{i'}} P_{i'}(\mathcal{M}'_{CSBVCG}(\mathcal{I}) \mid v_{-i'})$$

for each $v_{i'j}$ in Equation 10, and

$$\begin{aligned} \text{Equation 10} \quad &\leq \sum_{v_{i'j}:\ i'\neq\arg\max_{\hat{i}\neq i} v_{\hat{i}j}} \Pr_{\mathcal{D}_{i'j}}(v_{i'j}) \cdot \mathbb{E}_{v_{i'}\sim\mathcal{D}_{i'}} P_{i'}(\mathcal{M}'_{CSBVCG}(\mathcal{I}) \mid v_{-i'}) \\ &= \Pr_{v_{i'j}\sim\mathcal{D}_{i'j}}(i' \neq \arg\max_{\hat{i}\neq i} v_{\hat{i}j}) \cdot \mathbb{E}_{v_{i'}\sim\mathcal{D}_{i'}} P_{i'}(\mathcal{M}'_{CSBVCG}(\mathcal{I}) \mid v_{-i'}). \end{aligned} \tag{15}$$

Next, for each $v_{i'j}$ such that $i' = \arg\max_{\hat{i}\neq i} v_{\hat{i}j}$ and $i' \neq \arg\max_{\hat{i}\in N} v_{\hat{i}j}$, that is, those in Equation 11, i' is the potential winner of j under $BVCG(\mathcal{I}')$. Letting $M_{i'}$ be the potential winning set of i', we have that $BVCG(\mathcal{I}')$ makes a sale to i' if and only if

$$(v_{i'j} - \max_{\hat{i}\neq i,i'} v_{\hat{i}j}) + \sum_{k\in M_{i'}\setminus\{j\}} (v_{i'k} - \max_{\hat{i}\neq i,i'} v_{\hat{i}k}) \geq e^B_{i'}(\mathcal{D}_{i'}, v'_{-i'}),$$

that is, if and only if

$$\sum_{k\in M_{i'}\setminus\{j\}} (v_{i'k} - \max_{\hat{i}\neq i,i'} v_{\hat{i}k}) \geq e^B_{i'}(\mathcal{D}_{i'}, v'_{-i'}) - (v_{i'j} - \max_{\hat{i}\neq i,i'} v_{\hat{i}j}).$$

When this happens, the revenue generated from i' is

$$\begin{aligned} &\max_{\hat{i}\neq i,i'} v_{\hat{i}j} + \sum_{k\in M_{i'}\setminus\{j\}} \max_{\hat{i}\neq i,i'} v_{\hat{i}k} + e^B_{i'}(\mathcal{D}_{i'}, v'_{-i'}) \\ = \quad & v_{i'j} + \left[\sum_{k\in M_{i'}\setminus\{j\}} \max_{\hat{i}\neq i,i'} v_{\hat{i}k} + e^B_{i'}(\mathcal{D}_{i'}, v'_{-i'}) - (v_{i'j} - \max_{\hat{i}\neq i,i'} v_{\hat{i}j})\right], \end{aligned}$$

where the second part is exactly the revenue generated from selling $N\setminus\{j\}$ to i' with reserve prices $\overline{v_{-i'}}_{-j}$ and entry fee $e^B_{i'}(\mathcal{D}_{i'}, v'_{-i'}) - (v_{i'j} - \max_{\hat{i}\neq i,i'} v_{\hat{i}j})$. Accordingly, for each $v_{i'j}$ in Equation 11,

$$\begin{aligned} &\mathop{\mathbb{E}}_{(v_{i'j'})_{j'\neq j}\sim(\mathcal{D}_{i'j'})_{j'\neq j}} P_{i'}(BVCG(\mathcal{I}') \mid v'_{-i'}, v_{i'j}) \\ = \quad & v_{i'j} \cdot \Pr_{(v_{i'j'})_{j'\neq j}\sim(\mathcal{D}_{i'j'})_{j'\neq j}} (\text{A sale is made to } i' \text{ by } BVCG(\mathcal{I}')) \\ & + \mathop{\mathbb{E}}_{(v_{i'j'})_{j'\neq j}\sim(\mathcal{D}_{i'j'})_{j'\neq j}} P_{i'}(M_{i',\ N\setminus\{j\},\ \overline{v_{-i'}}_{-j},\ e^B_{i'}(\mathcal{D}_{i'},v'_{-i'})-(v_{i'j}-\max_{\hat{i}\neq i,i'} v_{\hat{i}j})}) \\ \leq \quad & v_{i'j} + \mathop{\mathbb{E}}_{(v_{i'j'})_{j'\neq j}\sim(\mathcal{D}_{i'j'})_{j'\neq j}} P_{i'}(M_{i',\ N\setminus\{j\},\ \overline{v_{-i'}}_{-j},\ e^B_{i'}(\mathcal{D}_{i'},v'_{-i'})-(v_{i'j}-\max_{\hat{i}\neq i,i'} v_{\hat{i}j})}) \\ \leq \quad & v_{i'j} + \mathop{\mathbb{E}}_{v_{i'}\sim\mathcal{D}_{i'}} P_{i'}(M_{i',\ N,\ \overline{v_{-i'}},\ e^B_{i'}(\mathcal{D}_{i'},v'_{-i'})-(v_{i'j}-\max_{\hat{i}\neq i,i'} v_{\hat{i}j})}) \\ \leq \quad & v_{i'j} + \mathop{\mathbb{E}}_{v_{i'}\sim\mathcal{D}_{i'}} P_{i'}(\mathcal{M}'_{CSBVCG}(\mathcal{I}) \mid v_{-i'}), \end{aligned}$$

where the first inequality is because the probability that a sale is made to i' is at most 1, the second is similar to Equation 13, and the third is similar to Equation 14. Thus,

$$\text{Equation 11} \leq \sum_{\substack{v_{i'j}:\ i'=\arg\max_{\hat{i}\neq i} v_{\hat{i}j},\\ i'\neq \arg\max_{\hat{i}\in N} v_{\hat{i}j}}} \Pr_{\mathcal{D}_{i'j}}(v_{i'j}) \cdot \left(v_{i'j} + \mathop{\mathbb{E}}_{v_{i'}\sim\mathcal{D}_{i'}} P_{i'}(\mathcal{M}'_{CSBVCG}(\mathcal{I}) \mid v_{-i'}) \right)$$

$$= \sum_{\substack{v_{i'j}:\ i'=\arg\max_{\hat{i}\neq i} v_{\hat{i}j},\\ i'\neq \arg\max_{\hat{i}\in N} v_{\hat{i}j}}} \Pr_{\mathcal{D}_{i'j}}(v_{i'j}) \cdot v_{i'j}$$

$$+ \Pr_{v_{i'j}\sim\mathcal{D}_{i'j}}\left(i' = \arg\max_{\hat{i}\neq i} v_{\hat{i}j},\ i' \neq \arg\max_{\hat{i}\in N} v_{\hat{i}j}\right) \cdot \mathop{\mathbb{E}}_{v_{i'}\sim\mathcal{D}_{i'}} P_{i'}(\mathcal{M}'_{CSBVCG}(\mathcal{I}) \mid v_{-i'}). \tag{16}$$

Moreover, for each $v_{i'j}$ such that $i' = \arg\max_{\hat{i}\in N} v_{\hat{i}j}$, that is, those in Equation 12, i' is the potential winner of j under $BVCG(\mathcal{I}')$. Again letting $M_{i'}$ be the potential winning set of i', we have that $BVCG(\mathcal{I}')$ makes a sale to i' if and only if

$$(v_{i'j} - \max_{\hat{i}\neq i,i'} v_{\hat{i}j}) + \sum_{k\in M_{i'}\setminus\{j\}} (v_{i'k} - \max_{\hat{i}\neq i,i'} v_{\hat{i}k}) \geq e^B_{i'}(\mathcal{D}_{i'}, v'_{-i'}).$$

That is, if and only if

$$(v_{i'j} - v_{ij}) + \sum_{k\in M_{i'}\setminus\{j\}} (v_{i'k} - \max_{\hat{i}\neq i,i'} v_{\hat{i}k}) \geq e^B_{i'}(\mathcal{D}_{i'}, v'_{-i'}) - v_{ij} + \max_{\hat{i}\neq i,i'} v_{\hat{i}j},$$

which happens exactly when selling N to i' with reserve prices $\overline{v_{-i'}}$ and entry fee $e^B_{i'}(\mathcal{D}_{i'}, v'_{-i'}) - v_{ij} + \max_{\hat{i}\neq i,i'} v_{\hat{i}j}$. Accordingly,

$$\begin{aligned}
&\text{Equation 12}\\
= &\sum_{v_{i'j}:\ i'=\arg\max_{\hat{i}\in N} v_{\hat{i}j}} \Pr_{\mathcal{D}_{i'j}}(v_{i'j}) \cdot \mathop{\mathbb{E}}_{(v_{i'j'})_{j'\neq j}\sim(\mathcal{D}_{i'j'})_{j'\neq j}} P_{i'}(BVCG(\mathcal{I}') \mid v'_{-i'}, v_{i'j})\\
= &\sum_{v_{i'j}:\ i'=\arg\max_{\hat{i}\in N} v_{\hat{i}j}} \Pr_{\mathcal{D}_{i'j}}(v_{i'j}) \cdot \mathop{\mathbb{E}}_{(v_{i'j'})_{j'\neq j}\sim(\mathcal{D}_{i'j'})_{j'\neq j}} P_{i'}(M_{i',\ N,\ \overline{v_{-i'}},\ e^B_{i'}(\mathcal{D}_{i'},v'_{-i'})-v_{ij}+\max_{\hat{i}\neq i,i'} v_{\hat{i}j}})\\
\leq &\sum_{v_{i'j}} \Pr_{\mathcal{D}_{i'j}}(v_{i'j}) \cdot \mathop{\mathbb{E}}_{(v_{i'j'})_{j'\neq j}\sim(\mathcal{D}_{i'j'})_{j'\neq j}} P_{i'}(M_{i',\ N,\ \overline{v_{-i'}},\ e^B_{i'}(\mathcal{D}_{i'},v'_{-i'})-v_{ij}+\max_{\hat{i}\neq i,i'} v_{\hat{i}j}})\\
= &\mathop{\mathbb{E}}_{v_{i'}\sim\mathcal{D}_{i'}} P_{i'}(M_{i',\ N,\ \overline{v_{-i'}},\ e^B_{i'}(\mathcal{D}_{i'},v'_{-i'})-v_{ij}+\max_{\hat{i}\neq i,i'} v_{\hat{i}j}})\\
\leq &\mathop{\mathbb{E}}_{v_{i'}\sim\mathcal{D}_{i'}} P_{i'}(\mathcal{M}'_{CSBVCG}(\mathcal{I}) \mid v_{-i'}),
\end{aligned} \tag{17}$$

where the first inequality is because we are counting in more events with non-negative revenue, and the second is again because $\mathcal{M}'_{CSBVCG}(\mathcal{I})$ computes the optimal entry fee to maximize the expected revenue generated from i'.

Note the difference between the upper bound for Equation 12 and those for Equations 10 and 11. In the previous two cases, the internal expectation is first compared with a mechanism that only sells $N \setminus \{j\}$ to i', and the latter can be directly upper bounded by a mechanism that sells N to i' using the same reserve prices for $N \setminus \{j\}$ and the same entry fee, which in turn is upper bounded by $\mathcal{M}'_{CSBVCG}$. For Equation 12, however, the internal expectation is directly compared with a mechanism that sells N to i', and this mechanism's expected revenue cannot be upper bounded by $\mathcal{M}'_{CSBVCG}$ unless the expectation is taken over $v_{i'}$ rather than $(v_{i'j'})_{j'\neq j}$.

Combining Equations 15, 16, 17, we have that, for each $v_{-i'}$ such that $v_{ij} > \max_{\hat{i}\neq i,i'} v_{\hat{i}j}$,

$$\begin{aligned}
& \mathbb{E}_{v_{i'}\sim\mathcal{D}_{i'}} P_{i'}(BVCG(\mathcal{I}') \mid v'_{-i'}) \\
\leq\ & \Pr_{v_{i'j}\sim\mathcal{D}_{i'j}}(i' \neq \arg\max_{\hat{i}\neq i} v_{\hat{i}j}) \cdot \mathbb{E}_{v_{i'}\sim\mathcal{D}_{i'}} P_{i'}(\mathcal{M}'_{CSBVCG}(\mathcal{I}) \mid v_{-i'}) \\
& + \sum_{\substack{v_{i'j}:\ i'=\arg\max_{\hat{i}\neq i} v_{\hat{i}j},\\ i'\neq \arg\max_{\hat{i}\in N} v_{\hat{i}j}}} \Pr_{\mathcal{D}_{i'j}}(v_{i'j}) \cdot v_{i'j} \\
& + \Pr_{v_{i'j}\sim\mathcal{D}_{i'j}}(i' = \arg\max_{\hat{i}\neq i} v_{\hat{i}j},\ i' \neq \arg\max_{\hat{i}\in N} v_{\hat{i}j}) \cdot \mathop{\mathbb{E}}_{v_{i'}\sim\mathcal{D}_{i'}} P_{i'}(\mathcal{M}'_{CSBVCG}(\mathcal{I}) \mid v_{-i'}) \\
& + \mathop{\mathbb{E}}_{v_{i'}\sim\mathcal{D}_{i'}} P_{i'}(\mathcal{M}'_{CSBVCG}(\mathcal{I}) \mid v_{-i'}) \\
\leq\ & 2 \mathop{\mathbb{E}}_{v_{i'}\sim\mathcal{D}_{i'}} P_{i'}(\mathcal{M}'_{CSBVCG}(\mathcal{I}) \mid v_{-i'}) + \sum_{\substack{v_{i'j}:\ i'=\arg\max_{\hat{i}\neq i} v_{\hat{i}j},\\ i'\neq \arg\max_{\hat{i}\in N} v_{\hat{i}j}}} \Pr_{\mathcal{D}_{i'j}}(v_{i'j}) \cdot v_{i'j},
\end{aligned} \tag{18}$$

where the second inequality is because

$$\Pr_{v_{i'j}\sim\mathcal{D}_{i'j}}(i' \neq \arg\max_{\hat{i}\neq i} v_{\hat{i}j}) + \Pr_{v_{i'j}\sim\mathcal{D}_{i'j}}(i' = \arg\max_{\hat{i}\neq i} v_{\hat{i}j},\ i' \neq \arg\max_{\hat{i}\in N} v_{\hat{i}j}) \leq 1.$$

Combining Equations 9 and 18, we have

$$\begin{aligned}
& \mathbb{E}_{v'\sim\mathcal{D}'} P_{i'}(BVCG(\mathcal{I}')) = \sum_{v_{-i'}} \Pr_{\mathcal{D}_{-i'}}(v_{-i'})\ \mathbb{E}_{v_{i'}\sim\mathcal{D}_{i'}} P_{i'}(BVCG(\mathcal{I}') \mid v'_{-i'}) \\
\leq\ & \sum_{v_{-i'}:\ v_{ij}\leq \max_{\hat{i}\neq i,i'} v_{\hat{i}j}} \Pr_{\mathcal{D}_{-i'}}(v_{-i'})\ \mathbb{E}_{v_{i'}\sim\mathcal{D}_{i'}} P_{i'}(\mathcal{M}'_{CSBVCG}(\mathcal{I}) \mid v_{-i'}) \\
& + \sum_{\substack{v_{-i'}:\\ v_{ij}>\max_{\hat{i}\neq i,i'} v_{\hat{i}j}}} \Pr_{\mathcal{D}_{-i'}}(v_{-i'}) \left[2 \mathop{\mathbb{E}}_{v_{i'}\sim\mathcal{D}_{i'}} P_{i'}(\mathcal{M}'_{CSBVCG}(\mathcal{I}) \mid v_{-i'}) + \sum_{\substack{v_{i'j}:\ i'=\arg\max_{\hat{i}\neq i} v_{\hat{i}j},\\ i'\neq \arg\max_{\hat{i}\in N} v_{\hat{i}j}}} \Pr_{\mathcal{D}_{i'j}}(v_{i'j}) \cdot v_{i'j} \right] \\
\leq\ & 2 \sum_{v_{-i'}} \Pr_{\mathcal{D}_{-i'}}(v_{-i'}) \mathop{\mathbb{E}}_{v_{i'}\sim\mathcal{D}_{i'}} P_{i'}(\mathcal{M}'_{CSBVCG}(\mathcal{I}) \mid v_{-i'}) \\
& + \sum_{\substack{v_{-i'}:\\ v_{ij}>\max_{\hat{i}\neq i,i'} v_{\hat{i}j}}} \Pr_{\mathcal{D}_{-i'}}(v_{-i'}) \sum_{\substack{v_{i'j}:\ i'=\arg\max_{\hat{i}\neq i} v_{\hat{i}j},\\ i'\neq \arg\max_{\hat{i}\in N} v_{\hat{i}j}}} \Pr_{\mathcal{D}_{i'j}}(v_{i'j}) \cdot v_{i'j} \\
=\ & 2 \mathop{\mathbb{E}}_{v\sim\mathcal{D}} P_{i'}(\mathcal{M}'_{CSBVCG}(\mathcal{I})) + \sum_{\substack{v:\ i=\arg\max_{\hat{i}} v_{\hat{i}j},\\ i'=\arg\max_{\hat{i}\neq i} v_{\hat{i}j}}} \Pr_{\mathcal{D}}(v) \cdot v_{i'j} = 2 \mathop{\mathbb{E}}_{v\sim\mathcal{D}} P_{i'}(\mathcal{M}'_{CSBVCG}(\mathcal{I})) + R_{i'}.
\end{aligned}$$

Accordingly,

$$\mathbb{E}_{v\sim\mathcal{D}} P_{i'}(\mathcal{M}'_{CSBVCG}(\mathcal{I})) \geq \frac{1}{2} \left(\mathbb{E}_{v'\sim\mathcal{D}'} P_{i'}(BVCG(\mathcal{I}')) - R_{i'} \right),$$

and Claim 3 holds. □

When there are more distributions unknown, by adding them back one by one, the revenue of $\mathcal{M}'_{CSBVCG}$ only gets larger. Theorem 8 holds by combining the revenue of $\mathcal{M}'_{CSBVCG}$ with the revenue of $\mathcal{M}_{CS1LA}$. □

References

[1] A. Archer and É. Tardos. Truthful mechanisms for one-parameter agents. In *42nd IEEE Symposium on Foundations of Computer Science (FOCS'01)*, pages 482–491, 2001.

[2] G. Artemov, T. Kunimoto, and R. Serrano. Robust virtual implementation: Toward a reinterpretation of the Wilson doctrine. *Journal of Economic Theory*, 148(2):424–447, 2013.

[3] R. J. Aumann. Agreeing to disagree. *The Annals of Statistics*, 4(6):1236–1239, 1976.

[4] P. Azar, J. Chen, and S. Micali. Crowdsourced Bayesian auctions. In *3rd Innovations in Theoretical Computer Science Conference (ITCS'12)*, pages 236–248, 2012.

[5] M. Babaioff, N. Immorlica, and R. Kleinberg. Matroids, secretary problems, and online mechanisms. In *8th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'07)*, pages 434–443, 2007.

[6] D. F. Bacon, D. C. Parkes, Y. Chen, M. Rao, I. Kash, and M. Sridharan. Predicting your own effort. In *11th International Conference on Autonomous Agents and Multigent Systems (AAMAS'12)*, pages 695–702, 2012.

[7] S. Baliga and R. Vohra. Market research and market design. *Advances in Theoretical Economics*, 3(1), 2003. Article 5.

[8] D. Bergemann and S. Morris. *Robust Mechanism Design.* World Scientific, Hackensack, NJ, 2012.

[9] D. Bergemann and S. Morris. Informational robustness and solution concepts. Princeton University William S. Dietrich II Economic Theory Center Research Paper No. 067_2014, `http://ssrn.com/abstract=2538677`, 2015.

[10] G. W. Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Weview*, 78(1):1–3, 1950.

[11] Y. Cai, C. Daskalakis, and S. M. Weinberg. An algorithmic characterization of multi-dimensional mechanisms. In *44th Annual ACM Symposium on Theory of Computing (STOC'12)*, pages 459–478, 2012.

[12] Y. Cai, C. Daskalakis, and S. M. Weinberg. Optimal multi-dimensional mechanism design: Reducing revenue to welfare maximization. In *53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS'12)*, pages 130–139, 2012.

[13] Y. Cai, N. Devanur, and S. M. Weinberg. A duality based unified approach to Bayesian mechanism design. In *48th Annual ACM Symposium on Theory of Computing (STOC'16),* to appear, 2016.

[14] S. Chawla, H. Fu, and A. Karlin. Approximate revenue maximization in interdependent value settings. In *15th ACM Conference on Economics and Computation (EC'15)*, pages 277–294, 2014.

[15] S. Chawla, J. D. Hartline, and R. Kleinberg. Algorithmic pricing via virtual valuations. In *8th ACM Conference on Electronic Commerce (EC'07)*, pages 243–251, 2007.

[16] S. Chawla, J. D. Hartline, D. L. Malec, and B. Sivan. Multi-parameter mechanism design and sequential posted pricing. In *43th ACM Symposium on Theory of Computing (STOC'10)*, pages 311–320, 2010.

[17] S. Chawla, D. L. Malec, and B. Sivan. The power of randomness in Bayesian optimal mechanism design. In *11th ACM Conference on Electronic Commerce (EC'10)*, pages 149–158, 2010.

[18] J. Chen and S. Micali. Mechanism design with possibilistic beliefs. *Journal of Economic Theory*, 156:77–102, 2013. Conference version appeared at *52nd IEEE Annual Symposium on Foundations of Computer Science (FOCS'11)*, pages 87–96, 2011.

[19] J. Chen, S. Micali, and R. Pass. Tight revenue bounds with possibilistic beliefs and level-k rationality. *Econometrica*, 83(4):1619–1639, 2015.

[20] R. Cole and T. Roughgarden. The sample complexity of revenue maximization. In *46th Annual ACM Symposium on Theory of Computing (STOC'14)*, pages 243–252, 2014.

[21] J. Cremer and R. P. McLean. Full extraction of the surplus in Bayesian and dominant strategy auctions. *Econometrica*, 56(6):1247–1257, 1988.

[22] N. R. Devanur and J. D. Hartline. Limited and online supply and the Bayesian foundations of prior-free mechanism design. In *10th ACM Conference on Electronic Commerce (EC'09)*, pages 41–50, 2009.

[23] N. R. Devanur, Z. Huang, and C. Psomas. The sample complexity of auctions with side information. In *48th Annual ACM Symposium on Theory of Computing (STOC'16),* to appear, 2016.

[24] P. Dhangwatnotai, T. Roughgarden, and Q. Yan. Revenue maximization with a single sample. *Games and Economic Behavior*, 91:318–333, 2015.

[25] S. Dobzinski, H. Fu, and R. D. Kleinberg. Optimal auctions with correlated bidders are easy. In *43th Annual ACM Symposium on Theory of Computing (STOC'11)*, pages 129–138, 2011.

[26] T. Gneiting and A. E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

[27] A. V. Goldberg, J. D. Hartline, A. R. Karlin, M. Saks, and A. Wright. Competitive auctions. *Games and Economic Behavior*, 55(2):242–269, 2006.

[28] J. C. Harsanyi. Games with incomplete information played by "Bayesian" players, I–III. Part I: The basic model. *Management Science*, 14(3):159–182, 1967.

[29] S. Hart and N. Nisan. Approximate revenue maximization with multiple items. In *13th ACM Conference on Electronic Commerce (EC'12)*, pages 656–656, 2012.

[30] J. Hartline and A. Karlin. Profit maximization in mechanism design. In N. Nisan, T. Roughgarden, É. Tardos, and V. V. Vazirani, editors, *Algorithmic Game Theory*, pages 331–361. Cambridge, 2007.

[31] J. D. Hartline and T. Roughgarden. Optimal mechanism design and money burning. In *40th Annual ACM Symposium on Theory of Computing (STOC'08)*, pages 75–84, 2008.

[32] J. D. Hartline and T. Roughgarden. Simple versus optimal mechanisms. In *10th ACM conference on Electronic commerce (EC'09)*, pages 225–234, 2009.

[33] Z. Huang, Y. Mansour, and T. Roughgarden. Making the most of your samples. In *16th ACM Conference on Economics and Computation (EC'15)*, pages 45–60, 2015.

[34] W. James. *Some Problems of Philosophy.* Harvard University Press, Cambridge, 1979.

[35] R. Kleinberg and S. M. Weinberg. Matroid prophet inequalities. In *44th Annual ACM Symposium on Theory of Computing (STOC'12)*, pages 123–136, 2012.

[36] D. Lehmann, L. I. Oćallaghan, and Y. Shoham. Truth revelation in approximately efficient combinatorial auctions. *Journal of the ACM*, 49(5):577–602, 2002.

[37] X. Li and A. C.-C. Yao. On revenue maximization for selling multiple independently distributed items. *Proceedings of the National Academy of Sciences*, 110(28):11232–11237, 2013.

[38] A. M. Manelli and D. R. Vincent. Multidimensional mechanism design: Revenue maximization and the multiple-good monopoly. *Journal of Economic Theory*, 137(1):153–185, 2007.

[39] M. R. Munafo, T. Pfeiffer, A. Altmejd, E. Heikensten, J. Almenberg, A. Bird, Y. Chen, B. Wilson, M. Johannesson, and A. Dreber. Using prediction markets to forecast research evaluations. *Royal Society Open Science*, 2(10):150–287, 2015.

[40] R. B. Myerson. Optimal auction design. *Mathematics of Operations Research*, 6(1):58–73, 1981.

[41] J. G. Riley and W. F. Samuelson. Optimal auctions. *The American Economic Review*, 71(3):381–392, 1981.

[42] A. Ronen. On approximating optimal auctions. In *3rd ACM Conference on Electronic Commerce (EC'01)*, pages 11–17, 2001.

[43] A. Rubinstein and S. M. Weinberg. Simple mechanisms for a subadditive buyer and applications to revenue monotonicity. In *16th ACM Conference on Economics and Computation (EC'15)*, pages 377–394, 2015.

[44] I. Segal. Optimal pricing mechanisms with unknown demand. *The American Economic Review*, 93(3):509–529, 2003.

[45] R. Wilson. Game-theoretic analysis of trading processes. In T. Bewley, editor, *Advances in Economic Theory: Fifth World Congress*, pages 33–77, 1987.

[46] L. Xia and V. Conitzer. Finite local consistency characterizes generalized scoring rules. In *21st International Joint Conference on Artificial Intelligence(IJCAI'09)*, pages 336–341, 2009.

[47] A. C.-C. Yao. An n-to-1 bidder reduction for multi-item auctions and its applications. In *26th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'15)*, pages 92–109, 2015.