From Bayesian to Crowdsourced Bayesian Auctions

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

November 7, 2016

Abstract

A strong assumption in Bayesian mechanism design is that the distributions of the players' private types are common knowledge to the designer and the players —the *common prior* assumption. An important problem that has received a lot of attention in both economics and computer science is to *repeatedly weaken this assumption in game theory* —the "Wilson's Doctrine". In this work we consider, for the first time in the literature, multi-item auctions where the knowledge about the players' value distributions is *scattered* among the players and the seller. Each one of them privately knows some or none of the value distributions, no constraint is imposed on who knows which distributions, and the seller does not know who knows what.

In such an unstructured information setting, we design mechanisms for unit-demand auctions and additive auctions, whose expected revenue approximates that of the optimal Bayesian mechanisms by "crowdsourcing" the players' and the seller's knowledge. Our mechanisms are *2-step dominant-strategy truthful* and the revenue *increases gracefully with the amount of knowledge the players have.* In particular, the revenue starts from *a constant fraction* of the revenue of the best known dominant-strategy truthful Bayesian mechanisms, and approaches *100 percent* of the later when the amount of knowledge increases. Our results greatly improve the literature on the relationship between the amount of knowledge in the system and what mechanism design can achieve. In some sense, our results show that the common prior assumption is without much loss of generality in Bayesian auctions if one is willing to give up a fraction of the revenue. To establish our results, for unit-demand auctions we prove a *projection lemma* that converts an important class of Bayesian mechanisms to crowdsourced Bayesian mechanisms. For additive auctions, we construct a *hybrid* of the Bayesian and the crowdsourced Bayesian settings for designing our mechanisms; and we use the *adjusted revenue*, a novel concept first considered by Yao (2015), to analyze the revenue of our mechanisms.

Keywords: mechanism design, reasoning about knowledge, information elicitation, multi-parameter auctions, dominant-strategy truthfulness

1 Introduction

Since the seminal work of [43] and [24], Bayesian auction design has been an extremely flourishing area in game theory. One of the main focuses is to generate revenue, by selling m heterogenous items to n players. Each player has, as his private information, a valuation function describing how much he values each subset of the items. The valuations are drawn from prior distributions. An important assumption in Bayesian mechanism design is that the distributions are commonly known by the seller and the players —the *common prior* assumption. In particular, Bayesian mechanisms have the distributions given as input. However, as pointed out by another seminal work [52], such common knowledge is "rarely present in experiments and never in practice", and "only by repeated weakening of common knowledge assumptions will the theory approximate reality."[1] To answer the fundamental question of what mechanism design can achieve without a common prior, an important line of work has been developed in the past two decades by both economists and computer scientists; see, for example, [35, 9, 2, 10, 20, 22]. We will discuss them more carefully in Section 1.2.

In this paper we design, for the first time in the literature, auction mechanisms that allow the knowledge about the players' valuation distributions to be *scattered among and privately owned by the players and the seller*, rather than commonly shared by them. More specifically, we focus on unit-demand auctions [17] and additive auctions [33] —two classes of auctions widely studied in the literature. A player's valuation function is specified by m values, one for each item. For each player i and item j, the value v_{ij} is independently drawn from a distribution $\mathcal{D}_{ij}$. Each player privately knows his own values and some (or none) of the distributions of the other players, like long-time competitors in the same market. There is no constraint about who knows which distributions: in particular, it is possible that nobody knows how the whole valuation profile is distributed, and some value distributions are not known by anybody. The seller may also know some of the distributions, but he does not know which player knows what. Interestingly, the intuition behind this model has long been considered by philosophers. In [38], the author discussed a world where "everything in the world might be known by somebody, yet not everything by the same knower."

Now that each player's knowledge is his private information, our mechanisms have neither the distributions nor the knowledge structure of the players as input, and must work by properly aggregating the private knowledge of the players (and the seller). Hence we refer to such mechanisms as *crowdsourced Bayesian mechanisms.* However, note that they are different from classical crowdsourcing practices, which aggregate human labor for tasks such as image labeling [51]. We formalize our model in Section 2. Due to lack of space, some preliminaries and most proofs are provided in the appendix. Below we briefly discuss our main results and techniques.

1.1 Main Results and Discussions

To keep the settings simple and highlight the main difficulties in the study, 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 appendix we show how to remove this assumption. Indeed, this assumption serves as a way for us to focus on the key problems in designing crowdsourced Bayesian auctions.

The solution concept and the revenue benchmark. Our mechanisms ask each player to (privately) report both his values and his knowledge to the seller. In order to incentivize a player to report his true knowledge about the other players, he must be able to conclude that the others will

[1] As also righteously put in [52], "game theory as we presently know it cannot proceed without the fulcrum of common knowledge." Thus, rather than rejecting the common prior assumption, the author of [52] raised the removal of this assumption as an important challenge for future studies.

report their true values —only then will their reported values distribute according to his knowledge. Accordingly, we design mechanisms that are *2-step dominant strategy truthful* (2-DST) [4]: no matter what knowledge the players may report about each other, it is *dominant* for each player 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. This is a natural extension of dominant-strategy truthfulness and a powerful solution concept under mutual knowledge of rationality.

Our goal is to design 2-DST crowdsourced Bayesian mechanisms whose expected revenue approximates the revenue of the optimal *Bayesian incentive compatible* (BIC) mechanism, denoted by OPT. We may also compare the revenue of our mechanisms with that of the optimal *dominant strategy truthful* (DST) Bayesian mechanism, denoted by OPT_D.[2]

When everything is known by somebody. Without a common prior, we use *knowledge graphs* to represent who knows what: the nodes are players and there is an edge from player i to player i' for item j if and only if i knows $\mathcal{D}_{i'j}$. When each player i has in-degree at least 1 for each item j, "everything is known by somebody". Letting $\tau_k = \frac{k}{(k+1)^{\frac{k+1}{k}}}$ for any $k \in [n-1]$, we have the following theorems, formalized in Section 3. Note that $\tau_1 = \frac{1}{4}$ and $\tau_k \to 1$ when k gets larger.

Theorems 1 and 2. (sketched) $\forall k \in [n-1]$, *when each distribution is known by at least k players, there is a 2-DST crowdsourced Bayesian mechanism for unit-demand auctions with revenue $\geq \frac{\tau_k}{24} \cdot OPT$, and such a mechanism for additive auctions with revenue $\geq \max\{\frac{1}{11}, \frac{\tau_k}{6+2\tau_k}\}OPT$.*

Theorem 3. (sketched) *When the knowledge graph is 2-connected,*[3] *there is a 2-DST crowdsourced Bayesian mechanism for single-good auctions with revenue $\geq (1-\frac{1}{n})OPT$.*

When not everything is known. If some distributions are not known by anybody, some players have in-degree 0 for some items in the knowledge graphs. In this case, it is easy to see that no crowdsourced Bayesian mechanism can be a bounded approximation to OPT. Instead, we consider a natural benchmark: the optimal BIC mechanism *applied to players and items for whom the distributions are known by somebody.* Note that when everything is known, this is exactly OPT. More precisely, letting $\mathcal{I}'$ be the Bayesian instance projected to the knowledge graphs and $OPT(\mathcal{I}')$ the optimal Bayesian revenue on $\mathcal{I}'$, we have the following theorems, formalized in Section 4.

Theorems 4 and 5. (sketched) *For any knowledge graphs, there is a 2-DST crowdsourced Bayesian mechanism for unit-demand auctions with revenue $\geq \frac{1}{96} \cdot OPT(\mathcal{I}')$, and such a mechanism for additive auctions with revenue $\geq \frac{1}{70} \cdot OPT(\mathcal{I}')$.*

The key questions in designing crowdsourced Bayesian mechanisms. It is not hard to see that a 2-DST crowdsourced Bayesian mechanism implies a DST Bayesian mechanism: simply ignore the reported knowledge and use the common prior instead. Therefore the revenue of 2-DST crowdsourced Bayesian mechanisms is upper-bounded by that of the optimal DST Bayesian mechanism. For this reason, the two key questions here are *which Bayesian mechanisms can be "crowdsourced"*, and *what is the relationship between the players' knowledge structure and the mechanisms' revenue.* The first question is directly related to the robustness of Bayesian mechanisms to the common prior assumption: if a Bayesian mechanism can be properly converted to a crowdsourced Bayesian mechanism without much loss of its revenue, then we have a strong evidence that the common prior is not that crucial for it to work. Furthermore, the second question aims at quantitative answers

[2] A Bayesian mechanism is BIC if it is a Bayesian Nash equilibrium for all players to report their true values, and DST if it is dominant for each player to report his true values.

[3] A directed graph is 2-connected if for any node i, the graph with i and all adjacent edges removed is still strongly connected. For knowledge graphs, this means no player is an "information hub", without whom the players will split into two parts such that one part has no information about the other.

about the connection between the players' knowledge and the revenue. Our theorems and their analysis provide a good understanding on these questions.

The difficulties and our techniques. Since crowdsourced mechanisms elicit the players' private knowledge about each other's value distributions, it is not surprising that they use *proper scoring rules* [11, 30] to reward the players for their reported knowledge, as in information elicitation [42, 45]. However, the role of scoring rules here is not as significant, because they do not help solving the main difficulties in designing crowdsourced mechanisms.

Indeed, the difficulties here are that, even without rewarding the players for their knowledge, (1) it is dominant for each player to report his true values, (2) reporting his true knowledge *never hurts him* and (3) the resulting revenue approximates the desired benchmark. When all three properties hold, by rewarding the players' knowledge using proper scoring rules in a straightforward manner, the mechanisms will be such that reporting his true knowledge is *strictly better than lying* for a player. In other words, without using scoring rules, each player's utility when reporting his true knowledge must be always greater than or equal to his utility when lying, and scoring rules are introduced solely to break the utility-ties and make it strictly better for a player to tell the truth.

Accordingly, when designing our mechanisms we focus on achieving these three properties. As the total reward is upper-bounded by a small constant, we may ignore it when (stating and) analyzing our theorems. For unit-demand auctions, we show in Section 3.1 a general *projection lemma*: roughly speaking, any Bayesian mechanism whose revenue is approximately optimal in the COPIES setting (i.e., each player is replaced by m copies, one for each item) [18] can be "crowdsourced". This immediately converts several important Bayesian mechanisms in the literature to crowdsourced mechanisms. For additive auctions, in Section 4 we design mechanisms that work under a *hybrid* of two Bayesian settings; and we show the *adjusted revenue* [54] serves as an important bridge between the Bayesian and the crowdsourced Bayesian revenue.

Conclusion and future directions. In this paper, we strictly weaken the common prior assumption and provide important insights about the relation between the amount of knowledge in the system and the achievable revenue. Since the common prior assumption implies every player has *correct and exact* knowledge about all distributions, in the main body of this paper we do not consider scenarios where the players have "insider" knowledge. If the insider knowledge is correct (i.e., is a refinement of the prior as in [4]), then our mechanisms' revenue *increases*; see Appendix F. If the insider knowledge may be wrong, the problem is closely related to *robust mechanism design* [9], which is a very important topic in game theory but not the focus of this paper.

Designing (approximately) optimal Bayesian mechanisms has been of great interest to computer scientists. A common prior is clean to work with, but it is well understood that this is a strong assumption. In some sense, our results show that *this 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.* An important future direction is to "crowdsource" not only DST but also BIC mechanisms. For example, the BIC mechanisms in [24, 12, 13] are optimal in their own settings, and it is unclear how to convert them to crowdsourced Bayesian mechanisms. As pointed out by [24], the common prior assumption seems to be crucial for its mechanism. Another important direction is to design crowdsourced Bayesian mechanisms that are efficient in terms of communication. Indeed, if the distributions are continuous or if their supports are exponentially large in m and n, then it may not be feasible for the players to report the distributions in their entirety. We would be interested in understanding what types of information about the distributions can be efficiently elicited from the players and how much revenue can be generated from them.

1.2 Additional Related Work

Bayesian auction design. In his seminal work [43], Myerson introduced the first optimal Bayesian mechanism for single-good auctions with independent values, which applies to many single-parameter settings [1]. Since then, there has been a huge literature on designing (approximately) optimal Bayesian mechanisms that are either DST or BIC, such as [24, 47, 48, 17, 36, 28, 16]; see [34] for an introduction to this literature. In particular, mechanisms for multi-parameter settings have been constructed recently. In [12, 13], the authors characterize optimal BIC mechanisms for combinatorial auctions. For unit-demand auctions, [19, 18, 39] construct DST Bayesian mechanisms that are constant approximations. For auctions with additive valuations, [33, 41, 6, 54] provide logarithmic or constant approximations under different conditions. [14] provides a duality-based approach and obtains better approximation ratios for both unit-demand auctions and additive auctions. Moreover, [49] studies revenue mechanisms with a single buyer and sub-additive valuations.

Removing the common prior assumption. Following [52], a lot of effort has been made in the literature trying to remove the common prior assumption in game theory [8, 50, 31]. In *prior-free mechanisms* [35, 25], the distribution is unknown and the seller learns about it from the values of randomly selected players. In [22, 27, 37, 26], the seller observes independent samples from the distribution before the auction begins. In [20, 21], the players have arbitrary possibilistic belief hierarchies about each other. In *robust mechanism design* [9, 2, 10], the players have arbitrary probabilistic belief hierarchies. Finally, an earlier work of the first author of this paper considers a preliminary model for crowdsourced Bayesian auctions [4]: rather than allowing the distributions to be arbitrarily scattered among the players and the seller, each player privately knows a refinement of *all* the distributions. Thus, although the model is indeed weaker than the common prior assumption, it is a very special case of our model which also allows the players to refine the prior, and the task is much easier there. Both the insights and the techniques are very different in the two papers.

Information elicitation. Using proper scoring rules [11, 23, 30] as an important tool, a lot of effort has been devoted to *information elicitation* [44, 42, 45, 55]. Here each player reports his private information and his knowledge about the distribution of the other players' private information, and is rewarded based on everybody's report. Different from auctions, there are no allocations or prices for items, and a player's utility is equal to his reward. Bayesian Nash equilibrium is a widely used solution concept in this literature. Many studies here rely on a common prior, but mechanisms without this assumption have also been investigated [53]. In some sense, our work can be considered as information elicitation in auctions without a common prior. Note that in information elicitation the players do not have any cost for revealing their knowledge. It would be interesting to include such costs in (their and) our model, to see how the mechanisms will change accordingly.

2 The Model

In unit-demand auctions and additive auctions, a player i's value for an item j, v_{ij}, is independently drawn from a distribution $\mathcal{D}_{ij}$. Let $\mathcal{D}_i = \mathcal{D}_{i1} \times \cdots \times \mathcal{D}_{im}$ and $\mathcal{D} = \mathcal{D}_1 \times \cdots \times \mathcal{D}_n$. All our settings are downward closed [36]. The players have quasi-linear utilities and are risk-neutral, as commonly considered in Bayesian auctions. We will use Brier's scoring rule [11], denoted by BSR, to reward the players' knowledge in our mechanisms. If $\mathcal{D}$ is a distribution over a state space Ω and $\omega \in \Omega$, then $BSR(\mathcal{D}, \omega) = 2\mathcal{D}(\omega) - ||\mathcal{D}||_2^2 + 1$. Note the range of BSR is $[0, 2]$. Following the discussions in Section 1, all our mechanisms continue to be "weakly" 2-DST without using scoring rules: it is just that reporting the true knowledge may not be *strictly* better than lying for a player. Accordingly, in the analysis we may ignore the reward part, which is upper-bounded by a small constant ϵ in the mechanisms. See Appendix A for more preliminaries.

Knowledge graphs. It is illustrative to consider the players' knowledge as directed graphs: there is a vector of *knowledge graphs*, $G = (G_1, \ldots, G_m)$, one for each item. Each G_j has n nodes, one for each player, and edge (i, i') is in G_j if and only if player i knows $\mathcal{D}_{i'j}$.[4] There is no constraint about the structure of the knowledge graphs: the same player's distributions for different items may be known by different players, and different players' distributions for the same item may also be known by different players. Each player knows his own out-going edges but not the whole graph.

We measure the amount of knowledge in the system by *the number of players knowing each distribution.* More precisely, 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. The vector G is k-bounded if all the knowledge graphs are k-bounded. Note that 1-boundedness is a very weak assumption about the players' knowledge, and in fact the weakest when "everything is known". On the other hand, the common prior assumption implies that all knowledge graphs are complete graphs, or $(n-1)$-bounded, which is the strongest assumption in our model. The seller's knowledge can be naturally incorporated into the knowledge graphs: consider him as a special "player 0", whose out-going edges represent the distributions he knows. All our mechanisms can naturally utilize the seller's knowledge, and we will not further discuss this issue.

Crowdsourced Bayesian mechanisms. Different from Bayesian mechanisms, a crowdsourced Bayesian mechanism has neither the distribution $\mathcal{D}$ nor the graph vector G as input. Instead, it asks each player i to report a valuation function b_i and a *knowledge* K_i, where $K_i = \times_{i' \neq i, j \in [m]} \mathcal{D}^i_{i'j}$ 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 corresponding distributions. K_i is i's *true knowledge* if $\mathcal{D}^i_{i'j} = \mathcal{D}_{i'j}$ whenever $(i, i') \in G_j$, and $\mathcal{D}^i_{i'j} = \perp$ otherwise. The mechanism maps a strategy profile $(b_i, K_i)_{i \in [n]}$ to an allocation and a price profile, which may be randomized.

Given the set $N = \{1, \ldots, n\}$ of players, the set $M = \{1, \ldots, m\}$ of items and the distribution $\mathcal{D}$, let $\hat{\mathcal{I}} = (N, M, \mathcal{D})$ be the Bayesian auction instance and $\mathcal{I} = (N, M, \mathcal{D}, G)$ a corresponding crowdsourced Bayesian instance, where G is a knowledge graph vector. To distinguish whether a mechanism $\mathcal{M}$ is a Bayesian or crowdsourced Bayesian mechanism, we explicitly write $\mathcal{M}(\hat{\mathcal{I}})$ or $\mathcal{M}(\mathcal{I})$. Again note that the latter does not mean $\mathcal{M}$ has $\mathcal{D}$ or G as input. The expected revenue of a mechanism $\mathcal{M}$ is denoted by $Rev(\mathcal{M})$, and sometimes by $\mathbb{E}_{\mathcal{D}} Rev(\mathcal{M})$ to emphasize the distribution.

The no-bluff assumption. As mentioned, we make the *no-bluff assumption* about the players in the main body of the paper and we show how to remove it in Appendix E. A player i in a crowdsourced Bayesian mechanism is *no-bluff* if, for any knowledge graph G_j and player i' with $(i, i') \notin G_j$, and for any strategy (b_i, K_i) of i, i reports $\perp$ for the corresponding distribution of i'. Note that for a player i' with $(i, i') \in G_j$, i may report any distribution about i', including $\perp$.

3 When Everything Is Known by Somebody

In this section we construct crowdsourced Bayesian mechanisms when "everything is known by somebody": that is, when G is k-bounded with $k \geq 1$. For any $k \in [n-1]$, let $\tau_k = \frac{k}{(k+1)^{\frac{k+1}{k}}}$. Note that τ_k is increasing in k, $\tau_1 = \frac{1}{4}$ and $\tau_k \to 1$.

[4] We could have defined the players' knowledge using the standard notion [32, 3, 29]: roughly speaking, the state space consists of all possible distributions of the valuation profile, and player i knows $\mathcal{D}_{i'j}$ if he is in an information set where all distributions have the (i', j)-th component equal to $\mathcal{D}_{i'j}$. However, the knowledge graph is a more succinct representation and is enough for the purpose of this paper.

3.1 Unit-demand Auctions

We start with a crowdsourced Bayesian mechanism $\mathcal{M}_{CSUD}$ for unit-demand auctions, which takes as input a positive number ϵ that can be arbitrarily small; see Mechanism 1. Intuitively, the mechanism randomly partitions the players into two groups, asks one group to report their knowledge about the other, and runs the DST Bayesian mechanism of [39], denoted by $\mathcal{M}_{UD}$, on the latter. However, we emphasize that this is not a black-box reduction from arbitrary Bayesian mechanisms. Instead, we prove a general *projection lemma* (Lemma 2) that allows such a reduction from an important class of Bayesian mechanisms, including $\mathcal{M}_{UD}$ which leads to the best approximation ratio. We have the following theorem, whose proof is provided in Appendix B.1 and sketched below.

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

1: Each player i reports to the seller 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. $q \in (0,1)$ 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 $\mathcal{D}^i_{i'j} \neq \perp$, then (1) set $\mathcal{D}'_{i'j} = \mathcal{D}^i_{i'j}$; (2) 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 (3) 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 $\mathcal{D}'_{ij} \equiv 0$ (i.e., 0 w.p. 1) and $b_{ij} = 0$.

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 $\mathcal{D}'_{ij} \equiv 0$.

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 1. *$\forall k \in [n-1]$, 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 $\mathbb{E}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{CSUD}(\mathcal{I})) \geq \frac{\tau_k}{24} \cdot OPT(\hat{\mathcal{I}}) - \epsilon$.*

From now on we may simply say that $\mathcal{M}_{CSUD}$ is a $\frac{\tau_k}{24}$-approximation.

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

Proof sketch. As mentioned before, part of the difficulties is to ensure that, even without rewarding the players for their knowledge (i.e., Step 5(2) in $\mathcal{M}_{CSUD}$), it is dominant for a player to report his true values, and reporting his true knowledge can never hurt him. In $\mathcal{M}_{CSUD}$, the use of the players' values and the use of their knowledge have been disentangled: for players in N_1, the mechanism only uses their knowledge but not their values; and the opposite holds for players in N_2.

For any player in N_2, whether he is in N_3 does not depend on his own strategy. Accordingly, given that mechanism $\mathcal{M}_{UD}$ is DST and each player is assigned to N_2 with positive probability, it is dominant for each player to report his true values in $\mathcal{M}_{CSUD}$, no matter what the reported knowledge profile $(K_1, \ldots, K_n)$ is. If a player ends up in N_1, then he is guaranteed to get no item and pay 0, thus reporting his true knowledge never hurts him. □

The difficulty in analyzing the revenue of $\mathcal{M}_{CSUD}$ is that, it runs the Bayesian mechanism on a smaller Bayesian instance: $\hat{\mathcal{I}}$ *projected* to the set of player-item pairs (i,j) such that $i \in N_3$ and $\mathcal{D}_{ij}$ has been reported. To understand how much revenue is lost by the projection, we consider the COPIES instance [18], $\hat{\mathcal{I}}^{CP} = (N^{CP}, M^{CP}, \mathcal{D}^{CP})$. It is obtained by replacing each player with m copies and each item with n copies, where a player i's copy j only wants item j's copy i, with the value distributed according to $\mathcal{D}_{ij}$. Thus $\hat{\mathcal{I}}^{CP}$ is a single-parameter auction, with $N^{CP} = N \times M$, $M^{CP} = M \times N$, and $\mathcal{D}^{CP} = \times_{(i,j) \in N^{CP}} \mathcal{D}_{ij}$. Now we lower-bound the optimal revenue of the *projected* COPIES instance. We elaborate the related definitions and the proofs in Appendix B.1. For now, for any $NM \subseteq N \times M$, let $\hat{\mathcal{I}}^{CP}_{NM}$ be $\hat{\mathcal{I}}^{CP}$ projected to NM, and denote by $OPT(\hat{\mathcal{I}}^{CP})_{NM}$ the revenue of the optimal Bayesian mechanism for $\hat{\mathcal{I}}^{CP}$ obtained from players in NM.

Lemma 2 (*The projection lemma*)**.** *For any $NM \subseteq N \times M$, $OPT(\hat{\mathcal{I}}^{CP}_{NM}) \geq OPT(\hat{\mathcal{I}}^{CP})_{NM}$.*

Proof sketch of Theorem 1. By [39] and [14], mechanism $\mathcal{M}_{UD}$ is a 6-approximation to $OPT(\hat{\mathcal{I}}^{CP})$, and the latter is at least $\frac{OPT}{4}$. Theorem 1 holds by combining these facts and Lemma 2, and by setting $q = 1 - (k+1)^{-\frac{1}{k}}$ in our mechanism. □

Corollary 1. *Let $\mathcal{M}$ be a DST Bayesian mechanism such that $\mathbb{E}\, Rev(\mathcal{M}(\hat{\mathcal{I}})) \geq \alpha OPT(\hat{\mathcal{I}}^{CP})$ for some $\alpha > 0$. When the knowledge graphs are k-bounded with $k \geq 1$, there exists a 2-DST crowdsourced Bayesian mechanism that uses $\mathcal{M}$ as a black-box and is a $\frac{\tau_k \alpha}{4}$-approximation to OPT.*

Remark. As k gets larger (although k can still be much smaller than n), the approximation ratio of $\mathcal{M}_{CSUD}$ converges to 24, the best known approximation to OPT by DST Bayesian mechanisms [14]. Moreover, by Lemma 5 of [18], $\mathcal{M}_{CSUD}$ is also a $\frac{\tau_k}{6}$-approximation to the optimal deterministic DST mechanism. Finally, by Corollary 1, the mechanisms in [17] and [18] automatically imply corresponding crowdsourced Bayesian mechanisms.

3.2 Additive Auctions

Next, we construct crowdsourced Bayesian mechanisms for additive auctions, which is simpler than and partly based on our mechanism for unit-demand auctions. More specifically, we start with three important Bayesian mechanisms: the *Individual Myerson* mechanism, denoted by IM, which sells each item separately using Myerson's mechanism [43]; the *Individual 1-Lookahead* mechanism, denoted by $\mathcal{M}_{1LA}$, which sells each item separately using the 1-Lookahead mechanism [48] and is a 2-approximation to Myerson's mechanism for each item; and the *Bundle VCG* mechanism [54, 14], denoted by $BVCG$, which is the VCG mechanism with an optimal entry fee for each player.

We construct three 2-DST crowdsourced Bayesian mechanism —$\mathcal{M}_{CSIM}$, $\mathcal{M}_{CSBVCG}$, and $\mathcal{M}_{CS1LA}$ respectively. Properly combining these gadgets together, we obtain a crowdsourced Bayesian mechanism for additive auctions, denoted by $\mathcal{M}_{CSA}$. The mechanisms are formally defined and analyzed in Appendix B.2. Below we state our theorem and briefly discuss the ideas.

Theorem 2. $\forall k \in [n-1]$*, 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 $\mathbb{E}_{\mathcal{D}} Rev(\mathcal{M}_{CSA}(\mathcal{I})) \geq \max\{\frac{1}{11}, \frac{\tau_k}{6+2\tau_k}\} \cdot OPT(\hat{\mathcal{I}}) - \epsilon$.*

Proof sketch. Mechanism $\mathcal{M}_{CSIM}$ is the simplest: on each item, it is similar to $\mathcal{M}_{CSUD}$ and uses random sampling to disentangle the use of the players' values and their knowledge. It is a τ_k-approximation to IM. Mechanism $\mathcal{M}_{CSBVCG}$ has a different structure and does not need random sampling: the $BVCG$ mechanism run on the knowledge reported by *all* players is automatically 2-DST, because the allocation and the price for a player i depend only on the reported values and the knowledge about him, not the distributions of the others. When everything is known by somebody, the revenue of $\mathcal{M}_{CSBVCG}$ matches that of $BVCG$. Mechanism $\mathcal{M}_{CS1LA}$ has a similar structure to $\mathcal{M}_{CSBVCG}$: the 1-Lookahead mechanism can also be run on the knowledge reported by all players.

By [14], $\max\{Rev(IM), Rev(BVCG)\} \geq \frac{OPT}{8}$. However, we cannot compute the crowdsourced mechanisms' expected revenue and choose the better one, because we do not know the distributions. Moreover, computing the expected revenue using the players' reported knowledge and then choosing the better one may cause the players to lie about their knowledge. Accordingly, we choose from $\mathcal{M}_{CSIM}$, $\mathcal{M}_{CS1LA}$ and $\mathcal{M}_{CSBVCG}$ randomly, with the probabilities depending on k. □

3.3 Single-Good Auctions

In the last part on crowdsourced auctions when "everything is known by somebody" we focus on single-good auctions. Recall that the optimal Bayesian mechanism according to Myerson [43] maps each player i's reported value b_i to the *(ironed) virtual value*, $\phi_i(b_i; \mathcal{D}_i)$. It then runs the second-price mechanism with reserve price 0 on the virtual values and maps the resulting "virtual price" back to the winner's value space, as his price. Following Section 3.2 and Lemma 5 in Appendix B.2, for any $k \geq 1$, when the knowledge graph is k-bounded, mechanism $\mathcal{M}_{CSIM}$ is a τ_k-approximation to Myerson's mechanism. However, we show that a different crowdsourced mechanism can be *nearly optimal* under an interesting knowledge structure.

Recall a directed graph is *strongly connected* if there is a path from any player i to any other player i'. Intuitively, in a knowledge graph, this means for any two players Alice and Bob, Alice knows a guy who knows a guy ... who knows Bob. Also recall a graph is *2-connected* if it remains strongly connected after removing any single node and the adjacent edges. In a knowledge graph this means there does not exist a crucial player as an "information hub", without whom the players will split into two parts, with one part having no information about the other. It is easy to see that strong connectedness and 2-connectedness respectively imply 1-boundedness and 2-boundedness, but not vice versa. In fact, a graph of n nodes can be $(\lfloor \frac{n}{2} \rfloor - 1)$-bounded without being connected.

When the knowledge graph is 2-connected, we consider the *Crowdsourced Myerson* mechanism $\mathcal{M}_{CSM}$ in Mechanism 2. To help understanding it, we illustrate in Figure 2 of Appendix B.3 the sets of players involved in the first round. We have the following theorem, proved in the appendix.

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

1: Each player i reports a value b_i and a knowledge $K_i = (\mathcal{D}^i_j)_{j \in N \setminus \{i\}}$.
2: Randomly choose a player a, let $S = \{j \mid \mathcal{D}^a_j \neq \perp\}$, $N' = N \setminus (\{a\} \cup S)$, 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 $p_i = p_i - \frac{\epsilon}{2n^2} \cdot BSR(\mathcal{D}^i_{i'}, b_{i'})$.

Theorem 3. *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 $\mathbb{E}_{\mathcal{D}} Rev(\mathcal{M}_{CSM}(\mathcal{I})) \geq (1 - \frac{1}{n}) OPT(\hat{\mathcal{I}}) - \epsilon$.*

Proof sketch. Note that $|S| \geq 2$ in Step 2 due to 2-connectedness. The mechanism again disentangles the use of the players' values and their knowledge. Indeed, when computing a player's virtual value in Step 10, his knowledge has not been used yet. If he is player i^* then his knowledge will not be used in the next round either. Only when a player is removed from S —that is, when it is guaranteed that he will not get the item, will his knowledge be used. This is why it never hurts a player to report his true knowledge. However, because player i^* is excluded from the set of "reporters", we need that there is still a reporter who knows a distribution for players in N': that is, there is an edge from $(N \setminus N') \setminus \{i^*\}$ to N' and player i^* is not an "information hub" between $N \setminus N'$ and N'. This is again guaranteed by 2-connectedness (note that strong connectedness alone is not enough). Accordingly, $\mathcal{M}_{CSM}$ does not stop until all players' distributions have been reported. Therefore $\mathcal{M}_{CSM}$ manages to run Myerson's mechanism after randomly excluding a player a. □

Remark. If the seller knows at least two distributions, the mechanism can use him as the starting point and achieve $OPT - \epsilon$. Since no crowdsourced Bayesian mechanism can be a $(\frac{1}{2} + \delta)$-approximation for any constant $\delta > 0$ when $n = 2$ [4], our result is tight. Interestingly, after obtaining our result, we have found that 2-connected graphs (though not as knowledge graphs) have been explored several times in the game theory literature [7, 46] for totally different problems.

4 When Not Everything Is Known

Finally, we consider the most general cases where the knowledge graphs can be totally arbitrary and may not even be 1-bounded: that is, some distributions may not be known by anybody. The mechanisms here can be directly applied to previous settings.

It is easy to see that, if for some player i and item j, nobody knows $\mathcal{D}_{ij}$, then no crowdsourced Bayesian mechanism can achieve a bounded approximation to OPT. Indeed, it is possible that all other values are constantly 0, and v_{ij} is the only source for revenue. Although a Bayesian mechanism can find the optimal reserve price given $\mathcal{D}_{ij}$, a crowdsourced Bayesian mechanism can only set the same reserve price for all distributions and the approximation ratio is unbounded.

Accordingly, when "not everything is known", we consider a natural benchmark which is the optimal Bayesian revenue *on players and items for whom the distributions are known in the crowdsourced setting.* More precisely, let $\hat{\mathcal{I}} = (N, M, \mathcal{D})$ be a Bayesian instance and $\mathcal{I} = (N, M, \mathcal{D}, G)$ a corresponding crowdsourced instance. Let $\mathcal{D}' = \times_{i \in N, j \in M} \mathcal{D}'_{ij}$ be such that $\mathcal{D}'_{ij} = \mathcal{D}_{ij}$ if there exists a player i' with $(i', i) \in G_j$, and $\mathcal{D}'_{ij}$ is constantly 0 otherwise; we refer to $\mathcal{D}'$ as *$\mathcal{D}$ projected on G.* Letting $\mathcal{I}' = (N, M, \mathcal{D}')$ be the resulting Bayesian instance, the benchmark is $OPT(\mathcal{I}')$. Note that this is a demanding benchmark to consider in crowdsourced settings: it takes into consideration all the knowledge processed by the players and, when everything is known by somebody, even if G is only 1-bounded, we have $\mathcal{I}' = \hat{\mathcal{I}}$ and $OPT(\mathcal{I}') = OPT$.

Unit-demand auctions. The difficulties in designing crowdsourced Bayesian mechanisms here turn out to be different from previous settings. Indeed, unit-demand auctions become easy: simply set $q = \frac{1}{2}$ in $\mathcal{M}_{CSUD}$ and keep all the rest unchanged. Thus we only state the theorem.

Theorem 4. *The revised mechanism $\mathcal{M}_{CSUD}$ for unit-demand auctions is 2-DST and, for any instances $\hat{\mathcal{I}} = (N, M, \mathcal{D})$, $\mathcal{I} = (N, M, \mathcal{D}, G)$ and $\mathcal{I}' = (N, M, \mathcal{D}')$ with $\mathcal{D}'$ being $\mathcal{D}$ projected on G, $\mathcal{M}_{CSUD}$ is a 96-approximation to $OPT(\mathcal{I}')$.*

Additive auctions. Additive auctions are now the hard part: the new mechanism $\mathcal{M}'_{CSA}$ is simple to construct but requires significant effort to analyze. Previously if any distribution about a player is not reported, mechanism $\mathcal{M}_{CSBVCG}$ does not sell his winning set (i.e., items for which he is the highest bidder) to anybody. Doing so ensures 2-DSTness, and all distributions are reported when everything is known. Unfortunately this may lead to 0 revenue when not everything is known.

To rectify the situation, we wish to *only throw away the player-item pairs for which the value distribution is unknown,* and apply $BVCG$ to the remaining. However, such a mechanism is not truthful. For example, if player i knows $\mathcal{D}_{-i}$ and the other players have reported $\mathcal{D}_i$ but nothing else, then it may be better for i not to report any knowledge, so that the mechanism throws away all the other players, giving him a low price. Instead, mechanism $\mathcal{M}'_{CSA}$ works on *a hybrid of the original Bayesian instance $\hat{\mathcal{I}}$ and the projected instance $\mathcal{I}'$.*

More specifically, $\mathcal{M}'_{CSA}$ is based on the *β-Bundling* mechanism [54], denoted by *Bund*. Since the mechanism is run on a hybrid of $\hat{\mathcal{I}}$ and $\mathcal{I}'$, the performance of $\mathcal{M}'_{CSA}$ cannot be lower-bounded by that of *Bund* on $\mathcal{I}'$, and must be *patched up with a second-price sale to compensate the revenue loss caused by throwing away some player-item pairs.* With $\mathcal{M}'_{CSA}$ defined in Mechanism 3, we have the following theorem, proved in Appendix C.

Mechanism 3. $\mathcal{M}'_{CSA}$

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 $M_i = \{j \mid i^*(j) = i\}$ be player i's winning set.
6: Partition M into M_i^1 and M_i^2 as follows: $\forall j \in M_i^1$, some i' has reported $\mathcal{D}^{i'}_{ij} \neq \perp$ (if there are more than one reporters, take the lexicographically first); and $\forall j \in M_i^2$, $\mathcal{D}^{i'}_{ij} = \perp$ for all i'.
7: $\forall j \in M_i^1$, set $\mathcal{D}'_{ij} = \mathcal{D}^{i'}_{ij}$; and $\forall j \in M_i^2$, set $\mathcal{D}'_{ij} \equiv 0$.
8: Compute the optimal entry fee e_i and reserve prices $(p'_j)_{j\in M_i^1}$ using *Bund* with input $(\mathcal{D}'_i, \beta_i)$, where $\beta_{ij} = \max_{i'\neq i} b_{i'j}$ $\forall j \in M$. If $e_i > 0$ then each $p'_j = \beta_{ij}$; if $e_i = 0$ then each $p'_j \geq \beta_{ij}$.
9: Sell $M_i^1 \cap M_i$ to player i according to *Bund*. That is, if $e_i > 0$ then do the following: if $\sum_{j\in M_i^1\cap M_i} b_{ij} \geq e_i + \sum_{j\in M_i^1\cap M_i} p'_j$, player 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. If $e_i = 0$ then do the following: for each item $j \in M_i^1 \cap M_i$, if $b_{ij} \geq p'_j$, player i gets item j with price p'_j; otherwise item j is not sold.
10: In addition, sell each item j in $M_i^2 \cap M_i$ to player i with price $p_j (= \beta_{ij})$.
11: **end for**

Theorem 5. *Mechanism $\mathcal{M}'_{CSA}$ is 2-DST for additive auctions and, for any instances $\hat{\mathcal{I}} = (N, M, \mathcal{D})$, $\mathcal{I} = (N, M, \mathcal{D}, G)$ and $\mathcal{I}' = (N, M, \mathcal{D}')$ with $\mathcal{D}'$ being $\mathcal{D}$ projected on G, $\mathcal{M}'_{CSA}$ is a 70-approximation to $OPT(\mathcal{I}')$.*

Proof sketch. In Steps 8 and 9, mechanism *Bund* decides how to sell to player i using $\mathcal{D}'_i$ and β_i, where the latter is determined by $b_{-i} \sim \mathcal{D}_{-i}$ under the players' truthful strategies: that is, $(\mathcal{D}'_i, \beta_i)$ is a *hybrid* of $\mathcal{I}'$ and $\hat{\mathcal{I}}$. Because for each item j, the highest bidder's distribution may not be reported and the reserve prices may be higher than in $\mathcal{I}'$, *Bund* may not be able to sell all the items that it would have sold in $\mathcal{I}'$. An important idea is to use the second-price sale (i.e., Step 10) to compensate the loss. However, the combined revenue still cannot be directly compared to that of *Bund* in $\mathcal{I}'$. In order to show that the second-price compensation is actually sufficient, the analysis consists of two major parts. First, we consider the optimal *β-adjusted revenue* [54] and show that it can be "hybridized": that is, the optimal adjusted revenue on the hybrid, together with the second-price sale, upper-bounds the optimal adjusted revenue on $\mathcal{I}'$; see Lemmas 12 and 14. Second, we show that the revenue of the 1-Lookahead mechanism [48] can be "hybridized" as well; see Lemma 13. □

Remark. By Theorem 5.1 of [54], $\mathcal{M}'_{CSA}$ is also a 69-approximation to OPT_D. Finally, by "hybridizing" the results of [14], it can be shown that $\mathcal{M}'_{CSA}$ is actually a 12-approximation to OPT. However, the analysis is far from being a black-box application and requires the hybridization of the counterpart of the β-adjusted revenue there.

Acknowledgement

The first author is grateful to Matt Weinberg for reading a draft of this paper and for the helpful discussions we had. The authors thank Constantinos Daskalakis, János Flesch, Hu Fu, Pinyan Lu, Silvio Micali, Rafael Pass, Andrés Perea, Elias Tsakas, several anonymous reviewers, and the participants of seminars at Stony Brook University, Shanghai Jiaotong University, Shanghai University of Finance and Economics, Maastricht University, MIT, and IBM Thomas J. Watson Research Center for their helpful comments. This work is partially supported by NSF CAREER Award No. 1553385.

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. In particular, 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 *true valuation profile* $v = (v_1, \ldots, v_n)$ 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 we take $\mathcal{D}_i = \mathcal{D}_{i1} \times \cdots \times \mathcal{D}_{im}$ for each player i and $\mathcal{D} = \mathcal{D}_1 \times \cdots \times \mathcal{D}_n$. 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* [36]: 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, 39] 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 [43] or single-minded auctions [40]. 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 $b = (b_1, \ldots, b_n)$. The *revenue* of a mechanism $\mathcal{M}$, denoted by $Rev(\mathcal{M})$, is the total price paid by the players. The expected revenue is denoted by $\mathbb{E}_{v \sim \mathcal{D}} Rev(\mathcal{M})$, or simply $Rev(\mathcal{M})$ when the distribution $\mathcal{D}$ is clear from the context.

Scoring rules are widely studied in economics and information elicitation. 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. Scoring rule 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 $\mathcal{D} \neq \mathcal{D}'$. Moreover, 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 [11]:

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

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 any $\mathcal{D}$ and ω.[5]

In a *crowdsourced Bayesian mechanism*, the distribution $\mathcal{D}$ is not given as input. Instead, a strategy s_i of player i consists of two parts, $s_i = (b_i, K_i)$, where b_i is a valuation function for i and K_i is a distribution for v_{-i}, representing i's knowledge about the other players.

Definition 1. *A crowdsourced Bayesian mechanism is* 2-step dominant strategy truthful *(2-DST) if the following two conditions hold:*

(1) For any player i, fixing any reported knowledge K_i, it is (weakly) dominant *for i to report his true valuation v_i. That is, for any K_i and any valuation function $b_i \neq v_i$,*

$$u_i((v_i, K_i), s_{-i}) \geq u_i((b_i, K_i), s_{-i})$$

for any strategy subprofile s_{-i}, and the inequality is strict for some s_{-i}.

(2) Given that all players report their true valuations, it is (weakly) dominant *for each player i to report his true knowledge. That is, denote i's true knowledge by $K_i = \times_{i'\neq i, j\in[m]}\mathcal{D}^i_{i'j}$, where $\mathcal{D}^i_{i'j} = \mathcal{D}_{i'j}$ if edge (i,i') is in the knowledge graph for item j and $\mathcal{D}^i_{i'j} = \perp$ otherwise. For any knowledge $K'_i \neq K_i$,*

$$u_i((v_i, K_i), (v_{-i}, K'_{-i})) \geq u_i((v_i, K'_i), (v_{-i}, K'_{-i}))$$

for any knowledge subprofile K'_{-i}, and the inequality is strict for some K'_{-i}.

It is worth pointing out that, without using proper scoring rules, it will be *very weakly dominant* for the players to report their true knowledge in our mechanisms: that is, reporting the truth never hurts a player and always leads to a utility greater than or equal to that of lying, but there may not be any case where it is strictly greater. If one takes the viewpoint that the players will tell the truth when they cannot benefit by lying, then there is no need to reward the players using scoring rules and the ϵ additive discount can be removed from our mechanisms. We keep them so as to be consistent with the literature of information elicitation, where the central problem is to buy information from the players.

B Proofs for Section 3

B.1 Proof of Theorem 1

Theorem 1. (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*

$$\mathbb{E}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{CSUD}(\mathcal{I})) \geq \frac{\tau_k}{24} \cdot OPT(\hat{\mathcal{I}}) - \epsilon.$$

We first prove the following lemma.

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

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

Proof. We start by showing the following claim.

Claim 1. *For any player i, knowledge K_i and true valuation $v_i = (v_{ij})_{j \in M}$, 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 valuation $b_i = (b_{ij})_{j \in M}$ for i and a strategy $s_{i'} = (b_{i'}, K_{i'})$ for each player $i' \neq i$. We need to compare $\mathbb{E}_{\mathcal{M}_{CSUD}} u_i(v_i, K_i)$ and $\mathbb{E}_{\mathcal{M}_{CSUD}} u_i(b_i, K_i)$, where the expectation is taken over the mechanism's random coins.[6]

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, and he gets the same utility for reporting any valuation. Thus $u_i(v_i, K_i) = u_i(b_i, K_i)$ in this case.

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 v_{ij}. Moreover, 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: that is $u_i(v_i, K_i) \geq u_i(b_i, K_i)$ and the inequality is strict for some strategy subprofiles s_{-i}.

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

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 $K_i = (\mathcal{D}^i_{i'j})_{i' \neq i, j \in M}$ as defined by the knowledge graphs: that is, for each item j, $\mathcal{D}^i_{i'j} = \mathcal{D}_{i'j}$ for all i' such that $(i, i') \in G_j$, and $\mathcal{D}^i_{i'j} = \perp$ otherwise.*

Proof. Arbitrarily fix a knowledge $K'_i = (\mathcal{D}'^i_{i'j})_{i' \neq i, j \in M}$ that is different from K_i. Under the no-bluff assumption, $\mathcal{D}'^i_{i'j} = \perp$ for all i' and j such that $(i, i') \notin G_j$. Moreover, arbitrarily fix a strategy $s_{i'}(v_{i'}) = (v_{i'}, K_{i'}(v_{i'}))$ for each player $i' \neq i$ and true valuation $v_{i'}$. Note that i''s reported valuation is always truthful, but his reported knowledge depends on his true valuation and need not be truthful. We now compare $\mathbb{E}_{\mathcal{M}_{CSUD}, \mathcal{D}_{-i}} u_i(v_i, K_i)$ and $\mathbb{E}_{\mathcal{M}_{CSUD}, \mathcal{D}_{-i}} u_i(v_i, K'_i)$, where the expectation is taken over both the mechanism's random coins and the distributions of the other players' true values.

Note that player i's reported knowledge does not matter if he is in N_2: that is

$$\mathbb{E}_{v_{-i} \sim \mathcal{D}_{-i}}[u_i(v_i, K_i) \mid i \in N_2] = \mathbb{E}_{v_{-i} \sim \mathcal{D}_{-i}}[u_i(v_i, K'_i) \mid i \in N_2].$$

We now focus on the case where he is in N_1. For any player $i' \in N_2$ and item $j \in M$ such that $(i, i') \in G_j$, whether i's report is used to define $\mathcal{D}'_{i'j}$ solely depends on the players in N_1 who are before him. If $\mathcal{D}'_{i'j}$ has not been defined when i is processed, then i's expected reward is maximized only by reporting $\mathcal{D}^i_{i'j} = \mathcal{D}_{i'j}$, because BSR is strictly proper. That is,

$$\mathbb{E}_{v_{-i} \sim \mathcal{D}_{-i}}[u_i(v_i, K_i) \mid i \in N_1] > \mathbb{E}_{v_{-i} \sim \mathcal{D}_{-i}}[u_i(v_i, K'_i) \mid i \in N_1].$$

Combining the two equations above, and because each player has a positive probability to be in N_1, we have

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

[6] Strictly speaking, each $s_{i'}$ should depend on $v_{i'}$ 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 $\mathcal{D}_{-i}$.

Thus it is dominant for i to report his true knowledge, and Claim 2 holds. $\square$

Lemma 1 follows directly from Claims 1 and 2. $\square$

Before analyzing the revenue of mechanism $\mathcal{M}_{CSUD}$, we first recall an important technique for reducing multi-parameter settings to single-parameter settings [18].

The COPIES setting. Given a unit-demand auction instance $\hat{\mathcal{I}} = (N, M, \mathcal{D})$, the corresponding COPIES auction instance is constructed as follows. We make m copies for each player, called *player copies*, and denote the resulting player set by $N^{CP} = N \times M$. We make n copies for each item, called *item copies*, and denote the resulting item set by $M^{CP} = M \times N$. Each player copy (i, j) has value $v_{ij} \sim \mathcal{D}_{ij}$ for the item copy (j, i), and 0 for all the other item copies.

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, and all other copies of i 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, with $\mathcal{D}^{CP} = \times_{(i,j)\in N^{CP}} \mathcal{D}_{ij}$. Intuitively, $\hat{\mathcal{I}}^{CP}$ involves more competition among the player copies: 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 revenue for COPIES is not too small compared with the optimal revenue for the original unit-demand instance, as shown by the following lemma.

Lemma 3. *[14] $OPT(\hat{\mathcal{I}}) \leq 4OPT(\hat{\mathcal{I}}^{CP})$.*

The projected setting. Next, we show that the optimal Bayesian mechanism for the COPIES setting satisfies the *projection lemma*. More specifically, let $NM \subseteq N \times M$ be a set of player-item pairs and N' be NM projected to N. The unit-demand instance $\hat{\mathcal{I}}_{NM} = (N', M, \mathcal{D}'_{N'})$, referred to as $\hat{\mathcal{I}}$ *projected to* NM, is such that $\mathcal{D}'_{ij} = \mathcal{D}_{ij}$ if $(i, j) \in NM$, and $\mathcal{D}'_{ij} \equiv 0$ otherwise. When NM can be represented as $N' \times M'$, we may also denote the projected instance by $\hat{\mathcal{I}}_{N',M'}$. If $N' = N$ or $M' = M$, we simplify the notation by using $\hat{\mathcal{I}}_{M'}$ or $\hat{\mathcal{I}}_{N'}$ instead.

Let $\hat{\mathcal{I}}^{CP}_{NM} = (N'^{CP}, M^{CP}, \mathcal{D}'^{CP}_{N'})$ be the COPIES instance corresponding to $\hat{\mathcal{I}}_{NM}$, also referred to as $\hat{\mathcal{I}}^{CP}$ *projected to* NM. That is, when projecting a COPIES instance to a set of player-item pairs, we still want the resulting instance to be a COPIES instance, thus we patch it up with the missing player-item pairs but with values constantly 0. The relations of these instances are illustrated in Figure 1.

$$
\begin{tikzcd}
\hat{\mathcal{I}} \arrow[r, "\text{COPIES}"] \arrow[d, "\text{projection}"'] & \hat{\mathcal{I}}^{CP} \arrow[d, "\text{projection}"] \\
\hat{\mathcal{I}}_{NM} \arrow[r, "\text{COPIES}"] & \hat{\mathcal{I}}^{CP}_{NM}
\end{tikzcd}
$$

Figure 1: Relations of the COPIES and the projected instances

Moreover, let $OPT(\hat{\mathcal{I}}^{CP})_{NM}$ be the revenue of the optimal Bayesian mechanism for $\hat{\mathcal{I}}^{CP}$ *projected to* NM: that is, for any pair $(i,j) \notin NM$, the price paid by the j-th copy of player i is not counted, even though this player copy may get the i-th copy of item j according to the optimal mechanism for $\hat{\mathcal{I}}^{CP}$. We have the following.

Lemma 2. (The projection lemma, restated) *For any* $NM \subseteq N \times M$,

$$\mathop{\mathbb{E}}_{\mathcal{D}'^{CP}_{N'}} OPT(\hat{\mathcal{I}}^{CP}_{NM}) \geq \mathop{\mathbb{E}}_{\mathcal{D}^{CP}} OPT(\hat{\mathcal{I}}^{CP})_{NM}.$$

Proof. Given the instance $\hat{\mathcal{I}}^{CP}_{NM}$, consider a Bayesian mechanism $\mathcal{M}'$ as follows:

- Patch up the instance to be exactly $\hat{\mathcal{I}}^{CP}$, including changing the distribution $\mathcal{D}'_{ij}$ to $\mathcal{D}_{ij}$ for any (i,j) with $i \in N'$ and $(i,j) \notin NM$;
- For any $(i,j) \in NM$, let i's copy j report his value, while for any $(i,j) \notin NM$, sample the value of i's copy j from $\mathcal{D}_{ij}$;
- Run the optimal DST Bayesian mechanism on $\hat{\mathcal{I}}^{CP}$, with the reported and sampled values;[7]
- Project the resulting outcome to NM.

The key is to show that $\mathcal{M}'$ is a DST Bayesian mechanism for $\hat{\mathcal{I}}^{CP}_{NM}$. Indeed, for any (i,j) such that $i \in N'$ and $(i,j) \notin NM$, the value of i's copy j in $\hat{\mathcal{I}}^{CP}_{NM}$ is 0 with probability 1, his reported value is not used by $\mathcal{M}'$, and at the end this player copy gets nothing and pays 0. Therefore it is dominant for this player copy to report his true value 0. For any $(i,j) \in NM$, the utility of i's copy j under $\mathcal{M}'$ is the same as that under the optimal DST Bayesian mechanism on $\hat{\mathcal{I}}^{CP}$. As it is dominant for this player copy to report his true value in the latter, so is it in $\mathcal{M}'$. Accordingly,

$$\mathop{\mathbb{E}}_{\mathcal{D}'^{CP}_{N'}} OPT(\hat{\mathcal{I}}^{CP}_{NM}) \geq \mathop{\mathbb{E}}_{\mathcal{D}'^{CP}_{N'}} Rev(\mathcal{M}'(\hat{\mathcal{I}}^{CP}_{NM})),$$

by the definition of OPT. By construction we have

$$\mathop{\mathbb{E}}_{\mathcal{D}'^{CP}_{N'}} Rev(\mathcal{M}'(\hat{\mathcal{I}}^{CP}_{NM})) = \mathop{\mathbb{E}}_{\mathcal{D}^{CP}} OPT(\hat{\mathcal{I}}^{CP})_{NM},$$

thus Lemma 2 holds. □

Notice that the projection lemma is not true for unit-demand settings in general, because the corresponding mechanism $\mathcal{M}'$ is not DST. Indeed, when the same player i has $(i,j) \in NM$ and $(i,j') \notin NM$ for some items j and j', he prefers receiving j to receiving j', even if the former leads to a smaller utility—his projected utility will be 0 under the latter.

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

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

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

[7]Since COPIES is a single-parameter setting, Myerson's mechanism is both the optimal BIC mechanism and the optimal DST mechanism here. In particular, the optimal BIC revenue equals the optimal DST revenue.

Notice that for any player i and item j,

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

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

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

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,

$$\Pr((i,j) \in NM_3) \geq \tau_k. \tag{1}$$

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 it is exactly $\hat{\mathcal{I}}$ projected to NM_3: that is, $\mathcal{D}'_{ij} = \mathcal{D}_{ij}$ for any $(i,j) \in NM_3$ and $\mathcal{D}'_{ij} \equiv 0$ 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 5. Since BSR is bounded in $[0,2]$, the reward each player i gets in this step is no more than $\frac{\epsilon}{n}$ and the total reward given to the players is no more than ϵ. Accordingly,

$$\begin{aligned}\mathop{\mathbb{E}}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{CSUD}(\mathcal{I})) &= \mathop{\mathbb{E}}_{NM_3} \mathop{\mathbb{E}}_{v_{NM_3}\sim\mathcal{D}_{NM_3}} Rev(\mathcal{M}_{UD}(\hat{\mathcal{I}}_{NM_3})) - R \\ &\geq \mathop{\mathbb{E}}_{NM_3} \mathop{\mathbb{E}}_{v_{NM_3}\sim\mathcal{D}_{NM_3}} Rev(\mathcal{M}_{UD}(\hat{\mathcal{I}}_{NM_3})) - \epsilon.\end{aligned} \tag{2}$$

Let $\hat{\mathcal{I}}^{CP}_{NM_3} = (N_3^{CP}, M^{CP}, \mathcal{D}'^{CP}_{N_3})$ be the COPIES instance corresponding to $\hat{\mathcal{I}}_{NM_3}$. By [39], there exists a Bayesian mechanism $\mathcal{M}^{CP}$ in the COPIES setting, which achieves less revenue than $\mathcal{M}_{UD}$ for the unit-demand instance, but is a 6-approximation to the optimal revenue for the COPIES instance. That is, for any NM_3,

$$\mathop{\mathbb{E}}_{v_{NM_3}\sim\mathcal{D}_{NM_3}} Rev(\mathcal{M}_{UD}(\hat{\mathcal{I}}_{NM_3})) \geq \mathop{\mathbb{E}}_{\mathcal{D}'^{CP}_{N_3}} Rev(\mathcal{M}^{CP}(\hat{\mathcal{I}}^{CP}_{NM_3})) \geq \frac{1}{6} \mathop{\mathbb{E}}_{\mathcal{D}'^{CP}_{N_3}} OPT(\hat{\mathcal{I}}^{CP}_{NM_3}). \tag{3}$$

Combining Inequalities 2 and 3, we have

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

By Lemma 2, we have

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

thus

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

Let $P_{ij}(OPT(\hat{\mathcal{I}}^{CP}))$ 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.

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

where the first equality is by the definition of the projection, the second is because sampling from $\mathcal{D}^{CP}$ is done independently from NM_3, and the third is 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 as follows.

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{\mathcal{D}^{CP}} \sum_{(i,j)\in N\times M} \Pr((i,j)\in NM_3)\cdot P_{ij}(OPT(\hat{\mathcal{I}}^{CP})) \\
\geq \ & \tau_k \mathop{\mathbb{E}}_{\mathcal{D}^{CP}} \sum_{(i,j)\in N\times M} P_{ij}(OPT(\hat{\mathcal{I}}^{CP})) = \tau_k OPT(\hat{\mathcal{I}}^{CP}) \geq \frac{\tau_k}{4} OPT(\hat{\mathcal{I}}), \qquad (7)
\end{aligned}
$$

where the first inequality is by Equation 1, 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

$$\mathbb{E}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{CSUD}(\mathcal{I})) \geq \frac{\tau_k}{24}\cdot OPT(\hat{\mathcal{I}}) - \epsilon,$$

and Theorem 1 holds. □

Remark. Note that Lemma 2 is only concerned with COPIES instances. Accordingly, for any Bayesian mechanism $\mathcal{M}$ whose revenue can be properly lower-bounded by the corresponding COPIES instance, similar analysis shows that $\mathcal{M}$ can be converted to a crowdsourced Bayesian mechanism. That is, the COPIES setting serves as a bridge between Bayesian and crowdsourced Bayesian mechanisms. Indeed, we have the following corollary, with the proof omitted.

Corollary 1. (restated) *Let $\mathcal{M}$ be a DST Bayesian mechanism such that, for some $\alpha > 0$ and for any Bayesian instance $\hat{\mathcal{I}}$,*

$$\mathop{\mathbb{E}}_{\mathcal{D}} Rev(\mathcal{M}(\hat{\mathcal{I}})) \geq \alpha \mathop{\mathbb{E}}_{\mathcal{D}^{CP}} OPT(\hat{\mathcal{I}}^{CP}).$$

When the knowledge graphs are k-bounded with $k \geq 1$, there exists a 2-DST crowdsourced Bayesian mechanism that uses $\mathcal{M}$ as a black-box and is a $\frac{\tau_k\alpha}{4}$-approximation to OPT.

B.2 Proof of Theorem 2

Theorem 2. (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*

$$\mathbb{E}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{CSA}(\mathcal{I})) \geq \max\{\frac{1}{11}, \frac{\tau_k}{6+2\tau_k}\}\cdot OPT(\hat{\mathcal{I}}) - \epsilon.$$

To prove Theorem 2, we start by introducing a simple mechanism $\mathcal{M}_{CSIM}$, which runs the mechanism $\mathcal{M}_{CSIM,j}$ (Mechanism 4) for each item j separately. Mechanism $\mathcal{M}_{CSIM,j}$ is similar to $\mathcal{M}_{CSUD}$, thus we have omitted many details. In particular, the following lemma is similar to Lemma 1 and we provide its statement only.

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 $q = 1 - (k+1)^{-\frac{1}{k}}$ and N_2 with probability $1 - q$.
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 $\hat{\mathcal{I}}_{N_3,j} = (N_3, \{j\}, \mathcal{D}'_{N_3,j})$ with the values being $(b_{ij})_{i \in N_3}$; and use the resulting allocation and prices to sell to players in N_3.

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 $\mathcal{M}_{CSIM}$.

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 1 and Lemma 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. □

Below we construct the crowdsourced 1-Lookahead mechanism $\mathcal{M}_{CS1LA}$ and show that its revenue matches that of mechanism $\mathcal{M}_{1LA}$ for any $k \geq 1$. Mechanism $\mathcal{M}_{CS1LA}$ runs the mechanism $\mathcal{M}_{CS1LA,j}$ (Mechanism 5) for each item j separately. Note that $\mathcal{M}_{CS1LA,j}$ does not partition the players into two groups. Also, it 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, 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. Moreover, 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. That is why the mechanism only tries to sell to the player with the highest value. We have the following two lemmas.

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 using 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, 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.

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 highest value among the other players if his distribution is not reported, and 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. $\mathbb{E}_{v \sim \mathcal{D}} Rev(\mathcal{M}_{CS1LA}(\mathcal{I})) \geq \mathbb{E}_{v \sim \mathcal{D}} Rev(\mathcal{M}_{1LA}(\hat{\mathcal{I}})) - \epsilon \geq \frac{1}{2}\mathbb{E}_{v \sim \mathcal{D}} Rev(IM(\hat{\mathcal{I}})) - \epsilon$.

Proof. When the players report their true values and true knowledge, despite of 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 mechanism $\mathcal{M}_{1LA}$ on the Bayesian instance $\hat{\mathcal{I}}_j = (N, \{j\}, \mathcal{D}_j)$. Accordingly,

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

where the second inequality is because the 1-Lookahead mechanism is a 2-approximation to the optimal Bayesian mechanism for each item j [48]. Thus Lemma 7 holds. □

Note that the approximation ratio of $\mathcal{M}_{CS1LA}$ does not depend on the specific value of k, as long as $k \geq 1$. Next, we describe the mechanism $\mathcal{M}_{CSBVCG}$, which approximates $BVCG$ in crowdsourced Bayesian settings.

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

- 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 $M_i = \{j \mid i^*(j) = i\}$ be player i's winning set.
  - If not all m distributions of i's values are reported, i gets no item and items in M_i are unsold.
  - 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 $BVCG$.
  - Sell M_i to player i according to $BVCG$. That is, 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.
- **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 the 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 $BVCG$: using the other players' highest reported value as the reserve price for each item, either player 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 [54], 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, nor the reserve prices for him. 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 the specific value of k, as long as $k \geq 1$. Indeed, notice the special structures of the two Bayesian mechanisms $\mathcal{M}_{1LA}$ and $BVCG$: the winning set of a player i solely depends on the players' values; the distribution $\mathcal{D}_i$ is only used to compute better reserve prices or entry fee to increase revenue; and the distribution $\mathcal{D}_{-i}$ is irrelevant to i. 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 such as Myerson's mechanism, all players' 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. When $k \leq 7$, it runs $\mathcal{M}_{CSBVCG}$ with probability $\frac{2}{11}$ and $\mathcal{M}_{CS1LA}$ with probability $\frac{9}{11}$; when $k > 7$, it runs $\mathcal{M}_{CSBVCG}$ with probability $\frac{\tau_k}{3+\tau_k}$ and $\mathcal{M}_{CSIM}$ with probability $\frac{3}{3+\tau_k}$. The choice of the two cases is to achieve the best approximation ratio for each k.

Proof of Theorem 2. 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 $k \leq 7$, we have $\frac{\tau_k}{6+2\tau_k} < \frac{1}{11}$. By Lemmas 7 and 9,

$$\begin{aligned}
&\mathbb{E}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{CSA}(\mathcal{I})) = \frac{2}{11}\mathbb{E}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{CSBVCG}(\mathcal{I})) + \frac{9}{11}\mathbb{E}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{CS1LA}(\mathcal{I})) \\
\geq\ & \frac{2}{11}\mathbb{E}_{v\sim\mathcal{D}} Rev(BVCG(\hat{\mathcal{I}})) + \frac{3}{11}\mathbb{E}_{v\sim\mathcal{D}} Rev(IM(\hat{\mathcal{I}})) + \frac{3}{11}\mathbb{E}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{1LA}(\hat{\mathcal{I}})) - \epsilon. \qquad (8)
\end{aligned}$$

When $k > 7$, we have $\frac{\tau_k}{6+2\tau_k} > \frac{1}{11}$. By Lemmas 5 and 9,

$$\begin{aligned}
\mathbb{E}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{CSA}(\mathcal{I})) &= \frac{\tau_k}{3+\tau_k}\mathbb{E}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{CSBVCG}(\mathcal{I})) + \frac{3}{3+\tau_k}\mathbb{E}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{CSIM}(\mathcal{I})) \\
&\geq \frac{\tau_k}{3+\tau_k}\mathbb{E}_{v\sim\mathcal{D}} Rev(BVCG(\hat{\mathcal{I}})) + \frac{3\tau_k}{3+\tau_k}\mathbb{E}_{v\sim\mathcal{D}} Rev(IM(\hat{\mathcal{I}})) - \epsilon. \qquad (9)
\end{aligned}$$

By [14], $OPT(\hat{\mathcal{I}}) \leq 2\mathbb{E}_{v\sim\mathcal{D}} Rev(BVCG(\hat{\mathcal{I}})) + 3\mathbb{E}_{v\sim\mathcal{D}} Rev(IM(\hat{\mathcal{I}})) + 3\mathbb{E}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{1LA}(\hat{\mathcal{I}}))$ $\leq 2\mathbb{E}_{v\sim\mathcal{D}} Rev(BVCG(\hat{\mathcal{I}})) + 6\mathbb{E}_{v\sim\mathcal{D}} Rev(IM(\hat{\mathcal{I}}))$. Combined with Inequalities 8 and 9, we have

$$\mathbb{E}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{CSA}(\mathcal{I})) \geq \max\left\{\frac{1}{11}, \frac{\tau_k}{6+2\tau_k}\right\} \cdot OPT(\hat{\mathcal{I}}) - \epsilon,$$

and Theorem 2 holds. □

B.3 Proof of Theorem 3

To help understanding mechanism $\mathcal{M}_{CSM}$, we illustrate in Figure 2 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 so far, drops all the other players from S, and adds the players whose distributions are reported for the first time by the dropped ones.

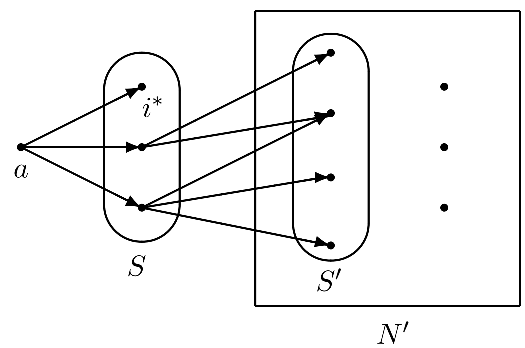

Figure 2: The sets of players involved in the first round of Mechanism $\mathcal{M}_{CSM}$.

Lemma 10. *$\mathcal{M}_{CSM}$ is 2-DST.*

Proof. The proof takes two steps.

Claim 3. *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.

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 the 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

(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_{-i} such that $\mathcal{D}'_i$ is defined, and the inequality is strict for some of them.

In sum, $\mathbb{E}_{\mathcal{M}_{CSM}} u_i(v_i, K_i) \geq \mathbb{E}_{\mathcal{M}_{CSM}} u_i(b_i, K_i)$ for all s_{-i} and the inequality is strict for some s_{-i}. Thus, fixing K_i in player i's strategy, it is dominant for i to report v_i and Claim 3 holds. □

Now we move to the second step.

Claim 4. *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 $K'_i = (\mathcal{D}'^i_j)_{j \neq i}$ that is different from K_i, and a strategy $s_j(v_j) = (v_j, K_j(v_j))$ for each $j \neq i$ and true value v_j. 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 Claim 3, 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 only 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). Same as Claim 3, $\mathcal{D}'_i$ 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 4 holds. □

Lemma 10 follows directly from Claims 3 and 4. □

Theorem 3. (restated) *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 $\mathbb{E}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{CSM}(\mathcal{I})) \geq (1-\frac{1}{n})OPT(\hat{\mathcal{I}}) - \epsilon$.*

Proof. It remains to show $\mathbb{E}_{v\sim\mathcal{D}} Rev(\mathcal{M}_{CSM}(\mathcal{I})) \geq (1-\frac{1}{n})OPT(\hat{\mathcal{I}}) - \epsilon$ under the players' truthful strategies, and the key is to explore the structure of the knowledge graph to make sure that the player with the highest virtual value is found by the mechanism with high probability.

More specifically, arbitrarily fix the player a chosen by the mechanism. Notice that throughout the mechanism, 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 $i \in (N \setminus N') \setminus S$, 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 edges going from $N \setminus N'$ to N' are from player i^*, and G becomes disconnected after 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 claim

$$\mathbb{E}_{v\sim\mathcal{D}}[Rev(\mathcal{M}_{CSM}(\mathcal{I}))|a] = OPT(\hat{\mathcal{I}}_a) - \epsilon. \tag{10}$$

To see why Equation 10 is true, note that by construction, in each round the mechanism keeps the player with the highest virtual value in S. Thus, the final player i^* 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 of Step 13 is the same as that of Myerson's mechanism on $\hat{\mathcal{I}}_a$, so is the revenue. It is easy to see that the total reward given to the players is no more than ϵ always. Therefore Equation 10 holds.

Finally, it remains to show that, by throwing away a random player a, the mechanism does not lose much revenue. For each player i, letting $P_i(OPT(\hat{\mathcal{I}}))$ be the expected price paid by i in Myerson's mechanism under $\hat{\mathcal{I}}$, we have $OPT(\hat{\mathcal{I}}) = \sum_{i\in N} P_i(OPT(\hat{\mathcal{I}}))$. Similar to the proof of Lemma 2, 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, thus it cannot generate more revenue than $OPT(\hat{\mathcal{I}}_a)$. As the expected revenue of $\mathcal{M}'$ is $\sum_{i\neq a} P_i(OPT(\hat{\mathcal{I}}))$, we have

$$OPT(\hat{\mathcal{I}}_a) \geq \mathbb{E}_{\mathcal{D}_{-a}} Rev(\mathcal{M}'(\hat{\mathcal{I}}_a)) = \sum_{i\neq a} P_i(OPT(\hat{\mathcal{I}})). \tag{11}$$

Combining Equations 10 and 11, we have

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

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

and Theorem 3 holds. $\square$

Remark. For additive auctions, when the knowledge graphs are 2-connected, instead of using mechanism $\mathcal{M}_{CS1LA}$ or $\mathcal{M}_{CSIM}$, one can use $\mathcal{M}_{CSM}$ for each item j. We thus have the following corollary, obtained by running $\mathcal{M}_{CSM}$ with probability $\frac{3}{4}$ and $\mathcal{M}_{CSBVCG}$ with probability $\frac{1}{4}$.

Corollary 2. *When the knowledge graphs are 2-connected, the revised mechanism $\mathcal{M}_{CSA}$ is 2-DST and $\underset{v\sim\mathcal{D}}{\mathbb{E}} Rev(\mathcal{M}_{CSA}(\mathcal{I})) \geq \frac{1}{8}(1-\frac{1}{n})OPT(\hat{\mathcal{I}}) - \epsilon$.*

C Proofs for Section 4

Theorem 5. (restated) *Mechanism $\mathcal{M}'_{CSA}$ is 2-DST for additive auctions and, for any instances $\hat{\mathcal{I}} = (N, M, \mathcal{D})$, $\mathcal{I} = (N, M, \mathcal{D}, G)$ and $\mathcal{I}' = (N, M, \mathcal{D}')$ with $\mathcal{D}'$ being $\mathcal{D}$ projected on G, $\mathcal{M}'_{CSA}$ is a 70-approximation to $OPT(\mathcal{I}')$.*

The following lemma is similar to its counterparts in previous sections and many details have been omitted.

Lemma 11. *$\mathcal{M}'_{CSA}$ is 2-DST.*

Proof. Note the difference between $\mathcal{M}_{CSBVCG}$ and $\mathcal{M}'_{CSA}$ is that, when a distribution about a player i's value for an item j is not reported, $\mathcal{M}_{CSBVCG}$ chooses not to sell anything to player i and charges him with price 0, while $\mathcal{M}'_{CSA}$ divides the items into two subsets M_i^1 and M_i^2, sells the first to i using mechanism *Bund*, and sells the second to i using the second-price mechanism.

Im mechanism $\mathcal{M}'_{CSA}$, given the distributions and values reported by the other players, M_i^1 and M_i^2 do not depend on player i's strategy, neither do the entry fee and reserve prices for M_i^1. Thus, it is dominant for i to report his true values for M_i^1, following the truthfulness of *Bund*; and it is dominant for him to report his true values for M_i^2, following the truthfulness of the second-price mechanism. Moreover, given that all players truthfully report their values, it is dominant for each i to report his true knowledge: indeed, the winning set M_i only depends on the players' reported values. Player i's reported knowledge does not affect the sets M_i^1 and M_i^2, or whether he gets some items or not, or the prices he pays. It only affects his reward under Brier's scoring rule. Thus $\mathcal{M}'_{CSA}$ is 2-DST and Lemma 11 holds. $\square$

In order to lower-bound the expected revenue of $\mathcal{M}'_{CSA}$, we first introduce several important concepts following [54]. For any single-player Bayesian instance $\hat{\mathcal{I}}_i = (\{i\}, M, \mathcal{D}_i)$ and any non-negative reserve-price vector $\beta_i = (\beta_{ij})_{j\in M}$, a single-player DST Bayesian mechanism is *β_i-exclusive* if it never sells an item j to i whenever his bid for j is no larger than β_{ij}. Denote by $Rev^X(\hat{\mathcal{I}}_i, \beta_i)$ the *optimal β_i-exclusive revenue* for $\hat{\mathcal{I}}_i$: that is, the superior over the revenue of β_i-exclusive mechanisms.

For any single-player DST Bayesian mechanism, its *β_i-adjusted revenue* on $\hat{\mathcal{I}}_i$ is its revenue minus *its social welfare generated from player-item pairs (i, j) such that i's bid for j is no larger than β_{ij}*. Denote by $Rev^A(\hat{\mathcal{I}}_i, \beta_i)$ the *optimal β_i-adjusted revenue* for $\hat{\mathcal{I}}_i$: that is, the superior over the β_i-adjusted revenue of all single-player DST Bayesian mechanisms. Note that if a mechanism is β_i-exclusive, then its β_i-adjusted revenue is exactly its revenue.

Given a Bayesian instance $\hat{\mathcal{I}} = (N, M, \mathcal{D})$, for each player i and a valuation subprofile $v_{-i} \sim \mathcal{D}_{-i}$, let $\beta_i(v_{-i})$ be such that $\beta_{ij}(v_{-i}) = \max_{i'\neq i} v_{i'j}$ for each item j. The *optimal β-adjusted revenue* for $\hat{\mathcal{I}}$ is

$$\mathbb{E}_{\mathcal{D}} Rev^A(\hat{\mathcal{I}}, \beta) = \sum_i \mathbb{E}_{v_{-i}\sim\mathcal{D}_{-i}} \mathbb{E}_{v_i\sim\mathcal{D}_i} Rev^A(\hat{\mathcal{I}}_i, \beta_i(v_{-i})),$$

obtained by applying the mechanism with the optimal $\beta_i(v_{-i})$-adjusted revenue for $\hat{\mathcal{I}}_i$ to sell to each player i, after seeing the valuation profile v. When v_{-i} is clear from the context, we may simply write β_i and β_{ij}.

Because we also consider the Bayesian instance $\mathcal{I}'$, let v' be v *projected on* G: that is, $v'_{ij} = v_{ij}$ if there exists a player i' with $(i', i) \in G_j$, and $v'_{ij} = 0$ otherwise. As v is distributed according to $\mathcal{D}$, v' is distributed according to $\mathcal{D}'$. Thus we sometimes directly sample $v' \sim \mathcal{D}'$ rather than sample v first and then map it to v'. Let $\beta'_{ij} = \max_{i' \neq i} v'_{i'j}$ for each player i and item j. The optimal β'-adjusted revenue for $\mathcal{I}'$ is defined respectively:

$$\mathbb{E}_{\mathcal{D}'} Rev^A(\mathcal{I}', \beta') = \sum_i \mathbb{E}_{v'_{-i} \sim \mathcal{D}'_{-i}} \mathbb{E}_{v'_i \sim \mathcal{D}'_i} Rev^A(\mathcal{I}'_i, \beta'_i(v'_{-i})).$$

It is important to emphasize that, given v_{-i} and $\beta_i(v_{-i})$, the optimal β_i-exclusive revenue and the optimal β_i-adjusted revenue are well defined for any Bayesian instance for player i, whether it is $\hat{\mathcal{I}}_i$ or $\mathcal{I}'_i$. In particular, we will consider $Rev^X(\mathcal{I}'_i, \beta_i)$ and $Rev^A(\mathcal{I}'_i, \beta_i)$, which are on the hybrid of $\hat{\mathcal{I}}_i$ and $\mathcal{I}'_i$. The optimal β-adjusted revenue on the hybrid of $\hat{\mathcal{I}}$ and $\mathcal{I}'$ are similarly defined:

$$\mathbb{E}_{\mathcal{D}} Rev^A(\mathcal{I}', \beta) = \sum_i \mathbb{E}_{v_{-i} \sim \mathcal{D}_{-i}} \mathbb{E}_{v'_i \sim \mathcal{D}'_i} Rev^A(\mathcal{I}'_i, \beta_i(v_{-i})).$$

To highlight that in the inner expectation, player i's value is v'_i even if it is obtained by first sampling $v_i \sim \mathcal{D}_i$ and then projecting on G, we may also write it as $\mathbb{E}_{v_i \sim \mathcal{D}_i} Rev^A(\mathcal{I}'_i, \beta_i; v'_i)$ and

$$\mathbb{E}_{\mathcal{D}} Rev^A(\mathcal{I}', \beta) = \sum_i \mathbb{E}_{v_{-i} \sim \mathcal{D}_{-i}} \mathbb{E}_{v_i \sim \mathcal{D}_i} Rev^A(\mathcal{I}'_i, \beta_i(v_{-i}); v'_i) = \mathop{\mathbb{E}}_{v \sim \mathcal{D}} \sum_i Rev^A(\mathcal{I}'_i, \beta_i; v'_i).$$

Finally, recall from Section 3.2 that mechanism IM is the Individual Myerson mechanism. Under the Bayesian instance $\mathcal{I}'$, $IM(\mathcal{I}')$ is the optimal revenue by selling each item separately. Since the total reward given by the scoring rule is at most ϵ, we will ignore it and treat $Rev(\mathcal{M}'_{CSA}(\mathcal{I}))$ as the revenue without giving out the rewards. Theorem 5 holds by the following two lemmas.

Lemma 12. $\mathop{\mathbb{E}}_{v \sim \mathcal{D}} Rev(\mathcal{M}'_{CSA}(\mathcal{I})) \geq \frac{1}{68} \mathop{\mathbb{E}}_{v' \sim \mathcal{D}'} Rev^A(\mathcal{I}', \beta')$.

Lemma 13. $\mathop{\mathbb{E}}_{v \sim \mathcal{D}} Rev(\mathcal{M}'_{CSA}(\mathcal{I})) \geq \frac{1}{2} \mathop{\mathbb{E}}_{v' \sim \mathcal{D}'} IM(\mathcal{I}')$.

Proof of Theorem 5. By Theorem 8.1 of [54], $\mathbb{E}_{v' \sim \mathcal{D}'} Rev^A(\mathcal{I}', \beta') + \mathop{\mathbb{E}}_{v' \sim \mathcal{D}'} IM(\mathcal{I}') \geq OPT(\mathcal{I}')$. Combining this inequality with Lemmas 12 and 13, we have

$$\mathop{\mathbb{E}}_{v \sim \mathcal{D}} Rev(\mathcal{M}'_{CSA}(\mathcal{I})) \geq \frac{1}{70} OPT(\mathcal{I}'),$$

and Theorem 5 holds. □

Below we prove the two lemmas.

Proof of Lemma 12. For each player i, the set M_i^1 of items for which i's distributions are reported and the set M_i^2 of items for which i's distributions are not reported are uniquely determined by the knowledge graph G in $\mathcal{I}$. Let $\mathcal{I}'_{i,M_i^1} = (\{i\}, M_i^1, \mathcal{D}'_{i,M_i^1})$ be the single-player Bayesian instance obtained by projecting $\mathcal{I}'$ on i and M_i^1, with $\mathcal{D}'_{i,M_i^1} = \times_{j\in M_i^1}\mathcal{D}'_{ij}$. We have

$$\begin{aligned}
& \mathbb{E}_{v\sim\mathcal{D}} Rev(\mathcal{M}'_{CSA}(\mathcal{I})) \\
= \quad & \mathbb{E}_{v\sim\mathcal{D}} \sum_i \left(Rev\left(Bund_i(\mathcal{D}'_i, v_{i,M_i^1}, \beta_i(v_{-i}))\right) + \sum_{j\in M_i^2\cap M_i} \max_{i'\neq i} v_{i'j} \right) \\
= \quad & \mathbb{E}_{v\sim\mathcal{D}} \sum_i \left(Rev\left(Bund_i(\mathcal{D}'_{i,M_i^1}, v_{i,M_i^1}, \beta_{i,M_i^1})\right) + \sum_{j\in M_i^2\cap M_i} \max_{i'\neq i} v_{i'j} \right), \qquad (12)
\end{aligned}$$

where $Bund_i$ is $Bund$ applied to player i in Steps 8 and 9. The first equality is by the construction of $\mathcal{M}'_{CSA}$. The second equality holds because in the execution of $Bund_i$ the items in M_i^2 do not affect anything, as player i's values for them are constantly 0 in $\mathcal{D}'_i$.

By Theorem 6.1 of [54], for any single-player Bayesian instance for a player i and any β_i, $Bund_i$ is an 8.5-approximation to the optimal β_i-exclusive revenue for this instance. Moreover, by Theorem 1 of [54], the optimal β_i-exclusive revenue is an 8-approximation to the optimal β_i-adjusted revenue for the same instance. Accordingly, we have

$$\begin{aligned}
& \mathop{\mathbb{E}}_{v\sim\mathcal{D}} \sum_i Rev(Bund_i(\mathcal{D}'_{i,M_i^1}, v_{i,M_i^1}, \beta_{i,M_i^1})) \\
= \quad & \sum_i \mathop{\mathbb{E}}_{v_{-i}\sim\mathcal{D}_{-i}} \mathop{\mathbb{E}}_{v_i\sim\mathcal{D}_i} Rev(Bund_i(\mathcal{D}'_{i,M_i^1}, v_{i,M_i^1}, \beta_{i,M_i^1})) \\
= \quad & \sum_i \mathop{\mathbb{E}}_{v_{-i}\sim\mathcal{D}_{-i}} \mathop{\mathbb{E}}_{v'_i\sim\mathcal{D}'_i} Rev(Bund_i(\mathcal{D}'_{i,M_i^1}, v'_{i,M_i^1}, \beta_{i,M_i^1})) \\
= \quad & \sum_i \mathop{\mathbb{E}}_{v_{-i}\sim\mathcal{D}_{-i}} \mathop{\mathbb{E}}_{v'_i\sim\mathcal{D}'_i} Rev(Bund_i(\mathcal{I}'_{i,M_i^1}, \beta_{i,M_i^1})) \\
\geq \quad & \frac{1}{8.5} \sum_i \mathop{\mathbb{E}}_{v_{-i}\sim\mathcal{D}_{-i}} \mathop{\mathbb{E}}_{v'_i\sim\mathcal{D}'_i} Rev^X(\mathcal{I}'_{i,M_i^1}, \beta_{i,M_i^1}) \\
= \quad & \frac{1}{8.5} \sum_i \mathop{\mathbb{E}}_{v_{-i}\sim\mathcal{D}_{-i}} \mathop{\mathbb{E}}_{v'_i\sim\mathcal{D}'_i} Rev^X(\mathcal{I}'_i, \beta_i) \\
\geq \quad & \frac{1}{8.5\times 8} \sum_i \mathop{\mathbb{E}}_{v_{-i}\sim\mathcal{D}_{-i}} \mathop{\mathbb{E}}_{v'_i\sim\mathcal{D}'_i} Rev^A(\mathcal{I}'_i, \beta_i) \\
= \quad & \frac{1}{68} \sum_i \mathop{\mathbb{E}}_{v_{-i}\sim\mathcal{D}_{-i}} \mathop{\mathbb{E}}_{v_i\sim\mathcal{D}_i} Rev^A(\mathcal{I}'_i, \beta_i; v'_i) \\
= \quad & \frac{1}{68} \mathop{\mathbb{E}}_{v\sim\mathcal{D}} \sum_i Rev^A(\mathcal{I}'_i, \beta_i; v'_i) = \frac{1}{68}\mathbb{E}_{v\sim\mathcal{D}} Rev^A(\mathcal{I}', \beta). \qquad (13)
\end{aligned}$$

The second equality above is because $\mathcal{D}'_i$ and $\mathcal{D}_i$ are the same when only M_i^1 is concerned, and $v'_{ij} = v_{ij}$ for all $j \in M_i^1$. The first inequality is by the relation between $Bund_i$ and Rev^X. The next equality is because i's values for items in M_i^2 are always 0 under $\mathcal{I}'_i$, thus even when these items are available, a β-exclusive mechanism does not sell them to i anyway. The second inequality is by the relation between Rev^X and the Rev^A. The next equality is because, as mentioned before, when sampling v_i from $\mathcal{D}_i$ and then projecting it on G, the resulting v'_i is distributed according to $\mathcal{D}'_i$.

Now we state the key lemma in our analysis, which connects the hybrid adjusted revenue with that for $\mathcal{I}'$. Recall that each β'_i is defined based on $v'_{-i} \sim \mathcal{D}'_{-i}$.

Lemma 14.

$$\mathbb{E}_{v\sim\mathcal{D}} Rev^A(\mathcal{I}', \beta) + \mathbb{E}_{v\sim\mathcal{D}} \sum_i \sum_{j\in M_i^2\cap M_i} \max_{i'\neq i} v_{i'j} \geq \mathbb{E}_{v'\sim\mathcal{D}'} Rev^A(\mathcal{I}', \beta').$$

Before proving Lemma 14, note that combining it with Equations 12 and 13 we have

$$\mathop{\mathbb{E}}_{v\sim\mathcal{D}} Rev(\mathcal{M}'_{CSA}(\mathcal{I})) \geq \frac{1}{68} \mathop{\mathbb{E}}_{v'\sim\mathcal{D}'} Rev^A(\mathcal{I}', \beta'). \tag{14}$$

Thus Lemma 12 holds. □

Next we prove Lemma 14.

Proof of Lemma 14. Arbitrarily fixing a player i and $v'_{-i} \sim \mathcal{D}'_{-i}$, let $\mathcal{M}^*$ be the single-player DST Bayesian mechanism with the optimal β'_i-adjusted revenue for $\mathcal{I}'_i$. Accordingly, $\mathcal{M}^*$ maximizes the following quantity:

$$\mathop{\mathbb{E}}_{v'_{-i}\sim\mathcal{D}'_{-i}} \mathop{\mathbb{E}}_{v'_i\sim\mathcal{D}'_i} [Rev(\mathcal{M}^*(\mathcal{I}'_i)) - \sum_{j:v'_{ij}\leq\beta'_{ij}} q_{ij} v'_{ij}],$$

where q_{ij} is the probability for i to get item j in $\mathcal{M}^*$ under v'_i. By definition,

$$\mathop{\mathbb{E}}_{v'\sim\mathcal{D}'} Rev^A(\mathcal{I}'_i, \beta'_i) = \mathop{\mathbb{E}}_{v'\sim\mathcal{D}'} [Rev(\mathcal{M}^*(\mathcal{I}'_i)) - \sum_{j:v'_{ij}\leq\beta'_{ij}} q_{ij} v'_{ij}]. \tag{15}$$

Again because sampling v from $\mathcal{D}$ and projecting it on G induces the same distribution for v' as $\mathcal{D}'$, we have

$$\mathop{\mathbb{E}}_{v'\sim\mathcal{D}'} [Rev(\mathcal{M}^*(\mathcal{I}'_i)) - \sum_{j:v'_{ij}\leq\beta'_{ij}} q_{ij} v'_{ij}] = \mathop{\mathbb{E}}_{v\sim\mathcal{D}} [Rev(\mathcal{M}^*(\mathcal{I}'_i; v'_i)) - \sum_{j:v'_{ij}\leq\beta'_{ij}} q_{ij} v'_{ij}], \tag{16}$$

where we explicitly include v'_i in the input of $\mathcal{M}^*$ to emphasize the projection.

Arbitrarily fix v_{-i} and let $\mathbb{U}_i = \{j | j \in M_i^1, \beta'_{ij} \neq \beta_{ij}\}$. It is clear that $\beta'_{ij} < \beta_{ij}$ for each $j \in \mathbb{U}_i$, as $v'_{-i} \leq v_{-i}$ (component-wise). For any v_i and the corresponding v'_i, as $v'_{ij} = 0$ for any $j \notin M_i^1$, we have

$$\begin{aligned}
\sum_{j:v'_{ij}\leq\beta'_{ij}} q_{ij} v'_{ij} &= \sum_{j\in M_i^1:v'_{ij}\leq\beta'_{ij}} q_{ij} v'_{ij} = \sum_{j\in M_i^1\setminus\mathbb{U}_i:v'_{ij}\leq\beta'_{ij}} q_{ij} v'_{ij} + \sum_{j\in\mathbb{U}_i:v'_{ij}\leq\beta'_{ij}} q_{ij} v'_{ij} \\
&= \sum_{j\in M_i^1\setminus\mathbb{U}_i:v'_{ij}\leq\beta_{ij}} q_{ij} v'_{ij} + \sum_{j\in\mathbb{U}_i:v'_{ij}\leq\beta_{ij}} q_{ij} v'_{ij} - \sum_{j\in\mathbb{U}_i:\beta'_{ij}<v'_{ij}\leq\beta_{ij}} q_{ij} v'_{ij} \\
&= \sum_{j\in M_i^1:v'_{ij}\leq\beta_{ij}} q_{ij} v'_{ij} - \sum_{j\in\mathbb{U}_i:\beta'_{ij}<v'_{ij}\leq\beta_{ij}} q_{ij} v'_{ij} \\
&= \sum_{j:v'_{ij}\leq\beta_{ij}} q_{ij} v'_{ij} - \sum_{j\in\mathbb{U}_i:\beta'_{ij}<v'_{ij}\leq\beta_{ij}} q_{ij} v'_{ij}.
\end{aligned} \tag{17}$$

Combining Equations 15, 16, 17 and taking summation over all players, we have

$$\begin{aligned}
& \mathbb{E}_{v'\sim\mathcal{D}'} Rev^A(\mathcal{I}', \beta') = \sum_i \mathop{\mathbb{E}}_{v'\sim\mathcal{D}'} Rev^A(\mathcal{I}'_i, \beta'_i) \\
= \ & \sum_i \mathop{\mathbb{E}}_{v\sim\mathcal{D}} [Rev(\mathcal{M}^*(\mathcal{I}'_i; v'_i)) - \sum_{j: v'_{ij} \le \beta'_{ij}} q_{ij} v'_{ij}] \\
= \ & \sum_i \mathop{\mathbb{E}}_{v\sim\mathcal{D}} \left(Rev(\mathcal{M}^*(\mathcal{I}'_i; v'_i)) - \sum_{j: v'_{ij} \le \beta_{ij}} q_{ij} v'_{ij} + \sum_{j\in\mathbb{U}_i: \beta'_{ij} < v'_{ij} \le \beta_{ij}} q_{ij} v'_{ij} \right) \\
\le \ & \sum_i \mathop{\mathbb{E}}_{v\sim\mathcal{D}} Rev^A(\mathcal{I}'_i, \beta_i; v'_i) + \sum_i \mathop{\mathbb{E}}_{v\sim\mathcal{D}} \sum_{j\in\mathbb{U}_i: \beta'_{ij} < v'_{ij} \le \beta_{ij}} q_{ij} v'_{ij} \\
= \ & \sum_i \mathop{\mathbb{E}}_{v\sim\mathcal{D}} Rev^A(\mathcal{I}'_i, \beta_i; v'_i) + \mathop{\mathbb{E}}_{v\sim\mathcal{D}} \sum_i \sum_{j\in\mathbb{U}_i: \beta'_{ij} < v_{ij} \le \beta_{ij}} q_{ij} v_{ij} \\
\le \ & \sum_i \mathop{\mathbb{E}}_{v\sim\mathcal{D}} Rev^A(\mathcal{I}'_i, \beta_i; v'_i) + \mathop{\mathbb{E}}_{v\sim\mathcal{D}} \sum_i \sum_{j\in\mathbb{U}_i: \beta'_{ij} < v_{ij} \le \beta_{ij}} v_{ij} \\
= \ & \mathbb{E}_{v\sim\mathcal{D}} Rev^A(\mathcal{I}', \beta) + \mathop{\mathbb{E}}_{v\sim\mathcal{D}} \sum_i \sum_{j\in\mathbb{U}_i: \beta'_{ij} < v_{ij} \le \beta_{ij}} v_{ij}. \qquad (18)
\end{aligned}$$

The first inequality above is because the first two terms in the expectation is exactly the β-adjusted revenue of mechanism $\mathcal{M}^*$ on $\mathcal{I}'_i$. The following equality holds because $v'_{ij} = v_{ij}$ for any $j \in M_i^1$. Finally, the second inequality is because $0 \le q_{ij} \le 1$ for any i, j.

Arbitrarily fix a valuation profile v and consider the term $\sum_i \sum_{j\in\mathbb{U}_i: \beta'_{ij} < v_{ij} \le \beta_{ij}} v_{ij}$ at the end of Equation 18. We show that each item j appears in the summation at most once. Indeed, for each v_{ij} that appears in the summation, $j \in M_i^1$ and $\mathcal{D}_{ij}$ is known in $\mathcal{I}$. As $v_{ij} > \beta'_{ij}$, player i has strictly higher value for j than any other player i' whose distribution for j is also known. For any such player i', $\beta'_{i'j} = v_{ij} > v_{i'j}$ and j is not in the set $\{j \in \mathbb{U}_{i'} : \beta'_{i'j} < v_{i'j} \le \beta_{i'j}\}$, which implies that $v_{i'j}$ does not appear in the summation. Moreover, for any player $i'' \neq i$ with $\mathcal{D}_{i''j}$ unknown, $j \notin M^1_{i''j}$ and $v_{i''j}$ does not appear in the summation either. Therefore item j only appears once in the summation. Accordingly,

$$\sum_i \sum_{j\in\mathbb{U}_i: \beta'_{ij} < v_{ij} \le \beta_{ij}} v_{ij} = \sum_{j:\ \exists!\ i \text{ s.t. } j\in\mathbb{U}_i,\ \beta'_{ij} < v_{ij} \le \beta_{ij}} v_{ij}. \qquad (19)$$

We now show that

$$\sum_{j:\ \exists!\ i \text{ s.t. } j\in\mathbb{U}_i,\ \beta'_{ij} < v_{ij} \le \beta_{ij}} v_{ij} \le \sum_i \sum_{j\in M_i^2 \cap M_i} \max_{i'\neq i} v_{i'j}, \qquad (20)$$

where the right-hand side is exactly the revenue generated by mechanism $\mathcal{M}_{CSA}$ in Step 10 and also the desired term in the statement of Lemma 14. Indeed, for each v_{ij} that appears in the left-hand side, because $v_{ij} \le \beta_{ij}$, player i's value for j is at most the second highest among all players. Letting $i^* = \arg\max_{i'\neq i} v_{i'j}$ with ties broken lexicographically, we have $\beta_{ij} = v_{i^*j}$ and v_{i^*j} is the highest value for j among all players. Since $\beta'_{ij} < \beta_{ij}$, $\mathcal{D}_{i^*j}$ is unknown and $j \in M^2_{i^*}$. Below we show $j \in M_{i^*}$, which then implies $j \in M^2_{i^*} \cap M_{i^*}$ and the price paid by i^* in Step 10 for item j is $\max_{i'\neq i^*} v_{i'j} \ge v_{ij}$.

When the distributions are generic, there are no ties in the players' values and v_{i^*j} is the unique maximum value for j, thus $j \in M_{i^*}$. For arbitrary distributions, problems occur when $v_{ij} = \beta_{ij}$ and $i < i^*$, which implies $j \notin M_{i^*}$. To deal with this special case, consider the following tie-breaking method for the players: while the value β_{ij} is still defined to be v_{i^*j}, we denote it by β_{ij}^+ if $i^* < i$ and β_{ij}^- if $i^* > i$. When $v_{ij} = \beta_{ij}$, we treat v_{ij} as strictly smaller if facing β_{ij}^+ and strictly larger if facing β_{ij}^-. All results proved in [54] and above continue to hold with respect to this tie-breaking method. Now for any v_{ij} that appears in the summation, either we have $v_{ij} < \beta_{ij}$ or we have $v_{ij} = \beta_{ij}$ and $i^* < i$, thus it is always the case that $j \in M_{i^*}$.

Accordingly,

$$\sum_{j:\ \exists!\ i \text{ s.t. } j \in \mathbb{U}_i,\ \beta'_{ij} < v_{ij} \leq \beta_{ij}} v_{ij} \leq \sum_{j:\ \exists!\ i^* \text{ s.t. } j \in M^2_{i^*} \cap M_{i^*}} \max_{i' \neq i^*} v_{i'j} = \sum_i \sum_{j \in M_i^2 \cap M_i} \max_{i' \neq i} v_{i'j},$$

and Equation 20 holds. Combining Equations 18, 19 and 20, we have

$$\mathbb{E}_{v' \sim \mathcal{D}'} Rev^A(\mathcal{I}', \beta') \leq \mathbb{E}_{v \sim \mathcal{D}} Rev^A(\mathcal{I}', \beta) + \mathop{\mathbb{E}}_{v \sim \mathcal{D}} \sum_i \sum_{j \in M_i^2 \cap M_i} \max_{i' \neq i} v_{i'j},$$

and Lemma 14 holds. □

We finish the whole analysis by proving Lemma 13.

Proof of Lemma 13. Similar to the proof of Lemma 12, the expected revenue generated by $Bund_i$ in Steps 8 and 9 is

$$\begin{aligned}
& \mathop{\mathbb{E}}_{v \sim \mathcal{D}} \sum_i Rev(Bund_i(\mathcal{D}'_i, v_{i,M_i^1}, \beta_i(v_{-i}))) \\
= & \sum_i \mathop{\mathbb{E}}_{v_{-i} \sim \mathcal{D}_{-i}} \mathop{\mathbb{E}}_{v_i \sim \mathcal{D}_i} Rev(Bund_i(\mathcal{D}_{i,M_i^1}, v_{i,M_i^1}, \beta_{i,M_i^1})) \\
\geq & \sum_i \mathop{\mathbb{E}}_{v_{-i} \sim \mathcal{D}_{-i}} \mathop{\mathbb{E}}_{v_i \sim \mathcal{D}_i} Rev(\mathcal{M}_{1LA}(\mathcal{D}_{i,M_i^1}, v_{i,M_i^1}, \beta_{i,M_i^1})) \\
= & \sum_i \mathop{\mathbb{E}}_{v \sim \mathcal{D}} Rev(\mathcal{M}_{1LA}(\hat{\mathcal{I}}_{i,M_i^1}, \beta_{i,M_i^1})).
\end{aligned} \tag{21}$$

The first equality above is again because items in M_i^2 does not affect $Bund_i$ given that player i's values for them are constantly 0 according to $\mathcal{D}'_i$, and $\mathcal{D}'_i$ and $\mathcal{D}_i$ coincide when only items in M_i^1 are concerned. The inequality holds because given $\mathcal{D}_{i,M_i^1}$ and β_{i,M_i^1}, mechanism $Bund_i$ chooses between optimally selling to i the items as a bundle (i.e., $e_i > 0$) and optimally selling to i each item separately (i.e., $e_i = 0$), whichever generates higher expected revenue over $v_{i,M_i^1} \sim \mathcal{D}_{i,M_i^1}$; while $\mathcal{M}_{1LA}$ is a particular mechanism that sells each item $j \in M_i^1$ to i separately based on $\mathcal{D}_{ij}$ and β_{ij}. Accordingly, letting $\mathcal{D}_j = \times(\mathcal{D}_{ij})_{i \in N}$ and $v_j \sim \mathcal{D}_j$ for each item j, we have

$$\begin{aligned}
& \mathbb{E}_{v \sim \mathcal{D}} Rev(\mathcal{M}'_{CSA}(\mathcal{I})) \\
= & \mathbb{E}_{v \sim \mathcal{D}} \sum_i \left(Rev\left(Bund_i(\mathcal{D}'_i, v_{i,M_i^1}, \beta_i(v_{-i})) \right) + \sum_{j \in M_i^2 \cap M_i} \max_{i' \neq i} v_{i'j} \right) \\
\geq & \sum_i \mathop{\mathbb{E}}_{v \sim \mathcal{D}} \left(Rev(\mathcal{M}_{1LA}(\hat{\mathcal{I}}_{i,M_i^1}, \beta_{i,M_i^1})) + \sum_{j \in M_i^2 \cap M_i} \max_{i' \neq i} v_{i'j} \right) \\
= & \sum_{j \in M} \sum_i \mathop{\mathbb{E}}_{v_j \sim \mathcal{D}_j} \left(\mathbf{I}_{j \in M_i^1} \cdot Rev(\mathcal{M}_{1LA}(\hat{\mathcal{I}}_{i,j}, \beta_{ij})) + \mathbf{I}_{j \in M_i^2 \cap M_i} \cdot \max_{i' \neq i} v_{i'j} \right).
\end{aligned} \tag{22}$$

Arbitrarily fixing an item j, we show that

$$\begin{aligned} & \sum_i \mathop{\mathbb{E}}_{v_j\sim\mathcal{D}_j}\left(\mathbf{I}_{j\in M_i^1}\cdot Rev(\mathcal{M}_{1LA}(\hat{\mathcal{I}}_{i,j},\beta_{ij}))+\mathbf{I}_{j\in M_i^2\cap M_i}\cdot\max_{i'\neq i} v_{i'j}\right) \\ \geq\ & \sum_i \mathop{\mathbb{E}}_{v'_j\sim\mathcal{D}'_j}\mathbf{I}_{j\in M_i^1}\cdot Rev(\mathcal{M}_{1LA}(\mathcal{I}'_{i,j},\beta'_{ij})). \end{aligned} \tag{23}$$

To do so, by definition, if $j\in M_i^1$ for a player i then in the Bayesian instance $\mathcal{I}'_{i,j}$ and given the reserve price β'_{ij}, the 1-Lookahead mechanism tries to sell j to i at price[8]

$$r'_{ij}=\arg\max_x x\Pr_{v'_{ij}\sim\mathcal{D}'_{ij}}(v'_{ij}\geq x|v'_{ij}\geq\beta'_{ij}).$$

Thus

$$\begin{aligned} & \sum_i \mathop{\mathbb{E}}_{v'_j\sim\mathcal{D}'_j}\mathbf{I}_{j\in M_i^1}\cdot Rev(\mathcal{M}_{1LA}(\mathcal{I}'_{i,j},\beta'_{ij})) \\ =\ & \sum_i \mathop{\mathbb{E}}_{v'_{-i,j}\sim\mathcal{D}'_{-i,j}}\mathop{\mathbb{E}}_{v'_{ij}\sim\mathcal{D}'_{ij}}\mathbf{I}_{j\in M_i^1}\cdot\mathbf{I}_{v'_{ij}\geq r'_{ij}}\cdot r'_{ij} \\ =\ & \sum_i \mathop{\mathbb{E}}_{v_{-i,j}\sim\mathcal{D}_{-i,j}}\mathop{\mathbb{E}}_{v_{ij}\sim\mathcal{D}_{ij}}\mathbf{I}_{j\in M_i^1}\cdot\mathbf{I}_{v'_{ij}\geq r'_{ij}(v'_{-i,j})}\cdot r'_{ij}(v'_{-i,j}), \end{aligned} \tag{24}$$

where the second equality is again by sampling v_j from $\mathcal{D}_j$ and projecting on G to get v'_j. We write r'_{ij} as $r'_{ij}(v'_{-i,j})$ to highlight this fact. Note that given $v'_{-i,j}$, $r'_{ij}(v'_{-i,j})$ depends on the distribution $\mathcal{D}'_{ij}$ but not on any concrete v'_{ij}. Also note that $r'_{ij}(v'_{-i,j})\geq\beta'_{ij}$. Next, we divide the last term in Equation 24 into two parts, depending on whether $r'_{ij}(v'_{-i,j})<\beta_{ij}$ or not. That is,

$$\begin{aligned} & \sum_i \mathop{\mathbb{E}}_{v_{-i,j}\sim\mathcal{D}_{-i,j}}\mathop{\mathbb{E}}_{v_{ij}\sim\mathcal{D}_{ij}}\mathbf{I}_{j\in M_i^1}\cdot\mathbf{I}_{v'_{ij}\geq r'_{ij}(v'_{-i,j})}\cdot r'_{ij}(v'_{-i,j}) \\ =\ & \sum_i \mathop{\mathbb{E}}_{v_{-i,j}\sim\mathcal{D}_{-i,j}}\mathop{\mathbb{E}}_{v_{ij}\sim\mathcal{D}_{ij}}\mathbf{I}_{j\in M_i^1}\cdot\mathbf{I}_{v'_{ij}\geq r'_{ij}(v'_{-i,j})}\cdot\left(\mathbf{I}_{r'_{ij}(v'_{-i,j})<\beta_{ij}}+\mathbf{I}_{r'_{ij}(v'_{-i,j})\geq\beta_{ij}}\right)\cdot r'_{ij}(v'_{-i,j}). \end{aligned} \tag{25}$$

For the $r'_{ij}(v'_{-i,j})<\beta_{ij}$ part in Equation 25, we have

$$\begin{aligned} & \sum_i \mathop{\mathbb{E}}_{v_{-i,j}\sim\mathcal{D}_{-i,j}}\mathop{\mathbb{E}}_{v_{ij}\sim\mathcal{D}_{ij}}\mathbf{I}_{j\in M_i^1}\cdot\mathbf{I}_{v'_{ij}\geq r'_{ij}(v'_{-i,j})}\cdot\mathbf{I}_{r'_{ij}(v'_{-i,j})<\beta_{ij}}\cdot r'_{ij}(v'_{-i,j}) \\ =\ & \sum_i \mathop{\mathbb{E}}_{v_{-i,j}\sim\mathcal{D}_{-i,j}}\mathop{\mathbb{E}}_{v_{ij}\sim\mathcal{D}_{ij}}\mathbf{I}_{j\in M_i^1}\cdot\left(\mathbf{I}_{v'_{ij}\geq\beta_{ij}>r'_{ij}(v'_{-i,j})}+\mathbf{I}_{\beta_{ij}>v'_{ij}\geq r'_{ij}(v'_{-i,j})}\right)\cdot r'_{ij}(v'_{-i,j}) \\ =\ & \sum_i \mathop{\mathbb{E}}_{v_{-i,j}\sim\mathcal{D}_{-i,j}}\mathop{\mathbb{E}}_{v_{ij}\sim\mathcal{D}_{ij}}\mathbf{I}_{j\in M_i^1}\cdot\left(\mathbf{I}_{v_{ij}\geq\beta_{ij}>r'_{ij}(v'_{-i,j})}+\mathbf{I}_{\beta_{ij}>v_{ij}\geq r'_{ij}(v'_{-i,j})}\right)\cdot r'_{ij}(v'_{-i,j}) \\ \leq\ & \sum_i \mathop{\mathbb{E}}_{v_{-i,j}\sim\mathcal{D}_{-i,j}}\mathop{\mathbb{E}}_{v_{ij}\sim\mathcal{D}_{ij}}\mathbf{I}_{j\in M_i^1}\cdot\left(\mathbf{I}_{v_{ij}\geq\beta_{ij}>r'_{ij}(v'_{-i,j})}\cdot\beta_{ij}+\mathbf{I}_{\beta_{ij}>v_{ij}\geq r'_{ij}(v'_{-i,j})}\cdot v_{ij}\right) \\ =\ & \sum_i \mathop{\mathbb{E}}_{v_{-i,j}\sim\mathcal{D}_{-i,j}}\mathop{\mathbb{E}}_{v_{ij}\sim\mathcal{D}_{ij}}\mathbf{I}_{j\in M_i^1}\cdot\mathbf{I}_{\beta_{ij}>r'_{ij}(v'_{-i,j})}\cdot\mathbf{I}_{v_{ij}\geq\beta_{ij}}\cdot\beta_{ij} \\ & \quad+\sum_i \mathop{\mathbb{E}}_{v_{-i,j}\sim\mathcal{D}_{-i,j}}\mathop{\mathbb{E}}_{v_{ij}\sim\mathcal{D}_{ij}}\mathbf{I}_{j\in M_i^1}\cdot\mathbf{I}_{\beta_{ij}>v_{ij}\geq r'_{ij}(v'_{-i,j})}\cdot v_{ij}. \end{aligned} \tag{26}$$

[8]Strictly speaking, under the Bayesian instance $\mathcal{I}'$, the Individual 1-Lookahead mechanism for item j works as follows. Given $v'_j\sim\mathcal{D}_j$, it finds the highest bidder i for j with ties broken lexicographically, as well as the second highest bid which is exactly β'_{ij}. It then tries to sell j to i at price r'_{ij}. For generic distributions, there are no ties in v'_j and the mechanism can be run on each instance $\mathcal{I}'_{ij}$ and β'_{ij} separately, without ever selling j to more than one players. For arbitrary distributions, this can be done by introducing proper tie-breaking rules.

The first equality above is by distinguishing whether $\beta_{ij} \le v'_{ij}$ or not. The second equality is because $v_{ij} = v'_{ij}$ whenever $j \in M_i^1$. The inequality is because $r'_{ij}(v'_{-i,j}) < \beta_{ij}$ following the indicator in the first term and $r'_{ij}(v'_{-i,j}) \le v_{ij}$ following the indicator in the second term. Finally, the last equality is because $\mathbf{I}_{v_{ij}\ge\beta_{ij}>r'_{ij}(v'_{-i,j})} = \mathbf{I}_{\beta_{ij}>r'_{ij}(v'_{-i,j})} \cdot \mathbf{I}_{v_{ij}\ge\beta_{ij}}$ in the first term.

For the first term in Equation 26, because the indicators $\mathbf{I}_{j\in M_i^1}$ and $\mathbf{I}_{\beta_{ij}>r'_{ij}(v'_{-i,j})}$ does not depend on v_{ij}, we have

$$\begin{aligned}
& \sum_i \mathop{\mathbb{E}}_{v_{-i,j}\sim\mathcal{D}_{-i,j}} \mathop{\mathbb{E}}_{v_{ij}\sim\mathcal{D}_{ij}} \mathbf{I}_{j\in M_i^1} \cdot \mathbf{I}_{\beta_{ij}>r'_{ij}(v'_{-i,j})} \cdot \mathbf{I}_{v_{ij}\ge\beta_{ij}} \cdot \beta_{ij} \\
= \; & \sum_i \mathop{\mathbb{E}}_{v_{-i,j}\sim\mathcal{D}_{-i,j}} \mathbf{I}_{j\in M_i^1} \cdot \mathbf{I}_{\beta_{ij}>r'_{ij}(v'_{-i,j})} \left(\mathop{\mathbb{E}}_{v_{ij}\sim\mathcal{D}_{ij}} \mathbf{I}_{v_{ij}\ge\beta_{ij}} \cdot \beta_{ij} \right) \\
\le \; & \sum_i \mathop{\mathbb{E}}_{v_{-i,j}\sim\mathcal{D}_{-i,j}} \mathbf{I}_{j\in M_i^1} \cdot \mathbf{I}_{\beta_{ij}>r'_{ij}(v'_{-i,j})} \left(\mathop{\mathbb{E}}_{v_{ij}\sim\mathcal{D}_{ij}} Rev(\mathcal{M}_{1LA}(\hat{\mathcal{I}}_{i,j}, \beta_{ij})) \right),
\end{aligned}$$

where the inequality is because on the Bayesian instance $\hat{\mathcal{I}}_{i,j}$ and given the reserve price β_{ij}, mechanism $\mathcal{M}_{1LA}$ chooses the optimal price r_{ij} to maximize the expected revenue and is no worse than simply setting $r_{ij} = \beta_{ij}$.

For the second term in Equation 26, for any valuation profile v, if there exists a player i is such that the two indicators both equal to 1, then we have $\beta_{ij} > v_{ij} \ge \beta'_{ij}$, because $\beta_{ij} > v_{ij} \ge r'_{ij}(v'_{-i,j})$ and $r'_{ij}(v'_{-i,j}) \ge \beta'_{ij}$. Thus the (lexicographically first) highest bidder i^* for j in v_j has his distribution unknown, and $j \in M_{i^*}^2 \cap M_{i^*}$; and player i is the (lexicographically first) highest bidder for j in v'_j, which is unique. Accordingly,

$$\begin{aligned}
& \sum_i \mathop{\mathbb{E}}_{v_{-i,j}\sim\mathcal{D}_{-i,j}} \mathop{\mathbb{E}}_{v_{ij}\sim\mathcal{D}_{ij}} \mathbf{I}_{j\in M_i^1} \cdot \mathbf{I}_{\beta_{ij}>v_{ij}\ge r'_{ij}(v'_{-i,j})} \cdot v_{ij} = \mathop{\mathbb{E}}_{v_j\sim\mathcal{D}_j} \sum_i \mathbf{I}_{j\in M_i^1} \cdot \mathbf{I}_{\beta_{ij}>v_{ij}\ge r'_{ij}(v'_{-i,j})} \cdot v_{ij} \\
\le \; & \mathop{\mathbb{E}}_{v_j\sim\mathcal{D}_j} \mathbf{I}_{j\in M_{i^*}^2\cap M_{i^*}} \cdot \max_{i:j\in M_i^1} v_{ij} \le \mathop{\mathbb{E}}_{v_j\sim\mathcal{D}_j} \mathbf{I}_{j\in M_{i^*}^2\cap M_{i^*}} \cdot \max_{i'\ne i^*} v_{i'j} = \mathop{\mathbb{E}}_{v_j\sim\mathcal{D}_j} \sum_i \mathbf{I}_{j\in M_i^2\cap M_i} \cdot \max_{i'\ne i} v_{i'j},
\end{aligned}$$

where the last equality is because i^* is the unique player such that the indicator is 1.

Combining the above two equations with Equation 26, for the $r'_{ij}(v'_{-i,j}) < \beta_{ij}$ part in Equation 25 we have

$$\begin{aligned}
& \sum_i \mathop{\mathbb{E}}_{v_{-i,j}\sim\mathcal{D}_{-i,j}} \mathop{\mathbb{E}}_{v_{ij}\sim\mathcal{D}_{ij}} \mathbf{I}_{j\in M_i^1} \cdot \mathbf{I}_{v'_{ij}\ge r'_{ij}(v'_{-i,j})} \cdot \mathbf{I}_{r'_{ij}(v'_{-i,j})<\beta_{ij}} \cdot r'_{ij}(v'_{-i,j}) \\
\le \; & \sum_i \mathop{\mathbb{E}}_{v_{-i,j}\sim\mathcal{D}_{-i,j}} \mathbf{I}_{j\in M_i^1} \cdot \mathbf{I}_{\beta_{ij}>r'_{ij}(v'_{-i,j})} \left(\mathop{\mathbb{E}}_{v_{ij}\sim\mathcal{D}_{ij}} Rev(\mathcal{M}_{1LA}(\hat{\mathcal{I}}_{i,j}, \beta_{ij})) \right) \\
& + \mathop{\mathbb{E}}_{v_j\sim\mathcal{D}_j} \sum_i \mathbf{I}_{j\in M_i^2\cap M_i} \cdot \max_{i'\ne i} v_{i'j} \\
= \; & \sum_i \mathop{\mathbb{E}}_{v_j\sim\mathcal{D}_j} \left(\mathbf{I}_{j\in M_i^1} \cdot \mathbf{I}_{\beta_{ij}>r'_{ij}(v'_{-i,j})} \cdot Rev(\mathcal{M}_{1LA}(\hat{\mathcal{I}}_{i,j}, \beta_{ij})) + \mathbf{I}_{j\in M_i^2\cap M_i} \cdot \max_{i'\ne i} v_{i'j} \right).
\end{aligned}$$

For the $r'_{ij}(v'_{-i,j}) \ge \beta_{ij}$ part in Equation 25, we immediately have $r'_{ij}(v'_{-i,j}) = r_{ij}(v_{-i,j})$ when the indicators are 1, because $\mathcal{D}'_{ij} = \mathcal{D}_{ij}$ and $v'_{ij} = v_{ij}$ whenever $j \in M_i^1$, and $r_{ij}(v_{-i,j})$ maximizes the expected revenue over $\mathcal{D}_{ij}$ conditional on $v_{ij} \ge \beta_{ij}$. Thus

$$
\begin{aligned}
& \sum_i \mathop{\mathbb{E}}_{v_{-i,j}\sim\mathcal{D}_{-i,j}} \mathop{\mathbb{E}}_{v_{ij}\sim\mathcal{D}_{ij}} \mathbf{I}_{j\in M_i^1}\cdot \mathbf{I}_{v'_{ij}\geq r'_{ij}(v'_{-i,j})}\cdot \mathbf{I}_{r'_{ij}(v'_{-i,j})\geq\beta_{ij}}\cdot r'_{ij}(v'_{-i,j}) \\
= & \sum_i \mathop{\mathbb{E}}_{v_{-i,j}\sim\mathcal{D}_{-i,j}} \mathop{\mathbb{E}}_{v_{ij}\sim\mathcal{D}_{ij}} \mathbf{I}_{j\in M_i^1}\cdot \mathbf{I}_{v_{ij}\geq r_{ij}(v_{-i,j})}\cdot \mathbf{I}_{r'_{ij}(v'_{-i,j})\geq\beta_{ij}}\cdot r_{ij}(v_{-i,j}) \\
= & \sum_i \mathop{\mathbb{E}}_{v_{-i,j}\sim\mathcal{D}_{-i,j}} \mathbf{I}_{j\in M_i^1}\cdot \mathbf{I}_{r'_{ij}(v'_{-i,j})\geq\beta_{ij}} \left(\mathop{\mathbb{E}}_{v_{ij}\sim\mathcal{D}_{ij}} \mathbf{I}_{v_{ij}\geq r_{ij}(v_{-i,j})}\cdot r_{ij}(v_{-i,j})\right) \\
= & \sum_i \mathop{\mathbb{E}}_{v_{-i,j}\sim\mathcal{D}_{-i,j}} \mathbf{I}_{j\in M_i^1}\cdot \mathbf{I}_{r'_{ij}(v'_{-i,j})\geq\beta_{ij}} \left(\mathop{\mathbb{E}}_{v_{ij}\sim\mathcal{D}_{ij}} Rev(\mathcal{M}_{1LA}(\hat{\mathcal{I}}_{i,j},\beta_{ij}))\right).
\end{aligned}
$$

Combining the above two equations with Equation 25 and then Equation 24, we have

$$
\begin{aligned}
& \sum_i \mathop{\mathbb{E}}_{v'_j\sim\mathcal{D}'_j} \mathbf{I}_{j\in M_i^1}\cdot Rev(\mathcal{M}_{1LA}(\mathcal{I}'_{i,j},\beta'_{ij})) \\
\leq & \sum_i \mathop{\mathbb{E}}_{v_j\sim\mathcal{D}_j} \left(\mathbf{I}_{j\in M_i^1}\cdot \mathbf{I}_{\beta_{ij}>r'_{ij}(v'_{-i,j})}\cdot Rev(\mathcal{M}_{1LA}(\hat{\mathcal{I}}_{i,j},\beta_{ij})) + \mathbf{I}_{j\in M_i^2\cap M_i}\cdot \max_{i'\neq i} v_{i'j}\right) \\
& + \sum_i \mathop{\mathbb{E}}_{v_{-i,j}\sim\mathcal{D}_{-i,j}} \mathbf{I}_{j\in M_i^1}\cdot \mathbf{I}_{r'_{ij}(v'_{-i,j})\geq\beta_{ij}} \left(\mathop{\mathbb{E}}_{v_{ij}\sim\mathcal{D}_{ij}} Rev(\mathcal{M}_{1LA}(\hat{\mathcal{I}}_{i,j},\beta_{ij}))\right) \\
= & \sum_i \mathop{\mathbb{E}}_{v_j\sim\mathcal{D}_j} \left(\mathbf{I}_{j\in M_i^1}\cdot Rev(\mathcal{M}_{1LA}(\hat{\mathcal{I}}_{i,j},\beta_{ij})) + \mathbf{I}_{j\in M_i^2\cap M_i}\cdot \max_{i'\neq i} v_{i'j}\right),
\end{aligned}
$$

and Equation 23 holds.

Taking summation over all items j on both sides of Equation 23 and combining with Equation 22, we have

$$
\begin{aligned}
& \mathbb{E}_{v\sim\mathcal{D}} Rev(\mathcal{M}'_{CSA}(\mathcal{I})) \geq \sum_{j\in M}\sum_i \mathop{\mathbb{E}}_{v'_j\sim\mathcal{D}'_j} \mathbf{I}_{j\in M_i^1}\cdot Rev(\mathcal{M}_{1LA}(\mathcal{I}'_{i,j},\beta'_{ij})) \\
= & \mathop{\mathbb{E}}_{v'\sim\mathcal{D}'} \sum_i \sum_{j\in M_i^1} Rev(\mathcal{M}_{1LA}(\mathcal{I}'_{i,j},\beta'_{ij})) = \mathop{\mathbb{E}}_{v'\sim\mathcal{D}'} Rev(\mathcal{M}_{1LA}(\mathcal{I}')) \geq \frac{1}{2}\mathop{\mathbb{E}}_{v'\sim\mathcal{D}'} IM(\mathcal{I}'),
\end{aligned}
$$

where the inequality is because $\mathcal{M}_{1LA}$ is a 2-approximation to Myerson's mechanism for each item. Thus Lemma 13 holds. □

D Crowdsourced Bayesian Auctions for Player-wise Information

In our main results, the knowledge graphs for different items can be totally different and there is no constraint about their structures. If all the knowledge graphs are the same, the problem becomes a special case which is also interesting: that is, a player knows another player's value distributions for all items. We denote this setting as *player-wise information*, in contrast with previous settings which can be considered as *(player, item)-wise information.*

Player-wise information setting can also model general combinatorial auctions. Recall that in such an auction, each player i's valuation function, v_i, maps each subset of items to a non-negative real, with $v_i(\emptyset) = 0$. Thus the distribution $\mathcal{D}_i$ is over such functions rather than values for individual items, and a player's values for two subsets of items can be arbitrarily correlated. Given such a

Bayesian instance $\hat{\mathcal{I}} = (N, M, \mathcal{D} = \times_{i\in N}\mathcal{D}_i)$, a corresponding crowdsourced Bayesian instance is denoted by $\mathcal{I} = (N, M, \mathcal{D}, G)$, where G is a single knowledge graph: an edge $(i, i') \in G$ means that player i knows $\mathcal{D}_{i'}$.

Optimal ϵ-BIC and BIC mechanisms have been designed in [12, 13] for combinatorial auctions, and little is known in the literature about how to construct (approximately) optimal DST Bayesian mechanisms here. In this section, *we provide simple blackbox reductions from Bayesian to crowdsourced Bayesian mechanisms,* showing that the latter can be obtained almost for free from the former when there is player-wise information. The mechanism $\mathcal{M}_{CSB}$ is defined in Mechanism 7, which takes a Bayesian mechanism $\mathcal{M}_B$ as a black-box. Notice that Bayesian Incentive Compatibility (BIC) are defined for crowdsourced Bayesian mechanisms in a similar way to that for Bayesian mechanisms: that is, all players reporting their true valuations and true knowledge is a Bayesian Nash equilibrium.

Mechanism 7. $\mathcal{M}_{CSB}$

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 $q = 1 - (k+1)^{-\frac{1}{k}}$ and N_2 with probability $1 - q$.
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 6. *For any $k \in [n-1]$, for any combinatorial 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_D), then $\mathcal{M}_{CSB}$ is a $\tau_k\sigma$-approximation to OPT (respectively, OPT_D).*

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. Moreover, if $\mathcal{M}_B$ is BIC, then $\mathcal{M}_{CSB}$ is also BIC almost 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 reporting the true knowledge 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.

The fact that $\mathcal{M}_{CSB}$ is a $\tau_k\sigma$-approximation whenever $\mathcal{M}_B$ is a σ-approximation can be proved in the same way as in the proof of Theorem 1, using similar notations. Below we only elaborate the case when $\mathcal{M}_B$ is a σ-approximation to OPT. More precisely, we need to show

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

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,

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

where the second inequality is again because, given that $\mathcal{M}_B$ is DST (respectively, BIC), running the corresponding optimal Bayesian mechanism on $\hat{\mathcal{I}}$ and projecting the outcome to N_3 is a DST (respectively, BIC) Bayesian mechanism on $\hat{\mathcal{I}}_{N_3}$. Thus Theorem 6 holds. □

Remark. With player-wise information, a player i's valuation distribution $\mathcal{D}_i$ is either "completely known" or "completely unknown". Thus for unit-demand auctions there is no need to adopt the COPIES setting to handle the scenario where only part of $\mathcal{D}_i$ is reported.

As before, the approximation ratio of $\mathcal{M}_{CSB}$ increases as k gets larger and converges to that of the Bayesian mechanism. When $k = 1$, it provides a 4-approximation using the optimal Bayesian mechanism as a black-box. Also note that the model studied in [4] is a very special case even compared with the player-wise information setting: that is, G is the complete graph and $k = n-1$. Since $\tau_{n-1} = \frac{n-1}{n^{n/(n-1)}} \to 1 - \frac{1}{n}$ when n gets larger, the revenue of our mechanism essentially matches that of [4] in this special case.

When not everything is known (that is, $k = 0$), we again consider the Bayesian instance projected to the knowledge graph, and benchmark against the optimal revenue there. The following corollary holds simply by setting $q = \frac{1}{2}$ in $\mathcal{M}_{CSB}$.

Corollary 3. *In the player-wise information setting, for any combinatorial 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 $\mathcal{M}_B$, the revised mechanism $\mathcal{M}_{CSB}$ is such that: if $\mathcal{M}_B$ is DST then $\mathcal{M}_{CSB}$ is 2-DST; 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_D), then $\mathcal{M}_{CSB}(\mathcal{I})$ is a $\frac{\sigma}{4}$-approximation to $OPT(\mathcal{I}')$ (respectively, $OPT_D(\mathcal{I}')$).*

E Removing the No-Bluff Assumption

To remove the no-bluff assumption, recall that, without this assumption, player i may report a distribution for another player i''s value for an item j even if $(i, i') \notin G_j$. However, if there exists a third player $\hat{i}$ who knows i''s distribution $\mathcal{D}_{i'j}$, and if player i is also rewarded for player $\hat{i}$'s report, then intuitively i has no incentive to bluff about i'. That is, *not only a player is paid for reporting the distributions he knows, but he is also paid for keeping quiet about the distributions he does not know and letting the experts speak.* Surely reporting $\mathcal{D}_{i'j}$ maximizes player i's expected reward, but he does not have the information to decide what $\mathcal{D}_{i'j}$ is.[9] Therefore, as long as reporting "$\perp$" gives

[9] Using the standard language, player i's information set contains at least two different distributions for i''s value for j.

player i the same utility as the unknown strategy of reporting $\mathcal{D}_{i'j}$, and as long as reporting any distribution other than $\mathcal{D}_{i'j}$ gives him a strictly smaller utility, player i will report "$\perp$" about $\mathcal{D}_{i'j}$.

More precisely, taking our mechanism for unit-demand auctions in Section 3 as an example, we can change the reward steps to the following:

- 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 proportional to r_{ij}.

Other parts of the mechanism remains 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. Thus the resulting crowdsourced mechanism is BIC. All the other mechanisms with knowledge graphs being at least 1-bounded can be changed similarly. We have the following theorem, whose proof has been omitted.

Theorem 7. *For any crowdsourced Bayesian mechanism $\mathcal{M}$ in previous sections with knowledge graphs being at least 1-bounded, there exists a BIC crowdsourced mechanism $\mathcal{M}'$ with the same expected revenue in the same auction settings, without making the no-bluff assumption. Moreover, all players reporting their true values and true knowledge is the unique Bayesian Nash equilibrium in $\mathcal{M}'$, besides the unachieved ones where each player reports the unknown true distributions.*

When not everything is known and the players may bluff, 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 the no-bluff assumption.

F Aggregating the Players' "Insider" Knowledge

As mentioned in Section 1, since the common prior assumption implies that every player has correct and exact (that is, no more, no less) knowledge about all the distributions, in the main body of this paper we do not consider scenarios where the players have "insider" knowledge. Incorrect insider knowledge has been studied in [9, 2, 20, 21, 10] and is not the focus of this paper. However, in some scenarios each player may have extra knowledge about the other players' value distributions besides the prior, and such knowledge is actually correct[10]: that is, it is a refinement of the prior. Different players' extra knowledge, although all correct, may refine the prior in different ways.

For example, when the prior distribution of a player i's value for an item j is uniform over $[0, 100]$, after v_{ij} is drawn another player i' may observe whether $v_{ij} \geq 50$ or not, and a third player i'' may observe whether $v_{ij} \in [20, 80]$. Thus player i' believes that v_{ij} is uniform over $[0, 50]$ or $(50, 100]$, depending on his signal; and player i'' also has his own beliefs depending on his signal.

The players' correct knowledge must be consistent with each other and, by combining their knowledge together one obtains an even better refinement of the prior. In the example above, if player i' observes $v_{ij} \leq 50$ and player i'' observes $v_{ij} \notin [20, 80]$, then it must be that $v_{ij} \in [0, 20)$ and the posterior distribution is uniform in this range. However, neither i' nor i'' knows this fact.

To model the players' extra knowledge, we equip the knowledge graphs in the crowdsourced Bayesian setting with *information sets*. To begin with, for any two players i, i' and item j such that $(i, i') \in G_j$, there is a partition $\mathcal{P}^i_{i'j}$ of the support of $\mathcal{D}_{i'j}$. After the true value $v_{i'j}$ is drawn, letting $S(v_{i'j})$ be the unique set in the partition that contains $v_{i'j}$, player i learns the fact that player i''s

[10] A similar scenario in the literature of contracts was considered in [15] with different concerns.

true value for j falls into $S(v_{i'j})$ and the posterior distribution is $\mathcal{D}_{i'j}|S(v_{i'j})$. More generally, the partitions may depend on player i's own true valuation v_i, because it is part of the information he has. To be consistent with the fact that different values are independently drawn, given v_i and the information sets observed by i for different distributions of the other players, i considers the posterior distributions to be independent.

It is not hard to see that for unit-demand auctions with arbitrary knowledge graphs, when the players have refined knowledge our mechanisms remain 2-DST, where a player's true knowledge is now the posterior distributions known by him. Since the optimal Bayesian revenue increases when the distributions are refined [4], the revenue of our mechanisms also increases. For additive auctions the same holds when everything is known, and a more careful analysis will be needed to understand how the revenue of our mechanism changes with respect to the players' refined knowledge when not everything is known. For single-good auctions when the knowledge-graph is 2-connected, one needs to compute a player's virtual value only when his distribution is reported for the first time, otherwise the players may not report their true knowledge.

Finally, although our mechanisms and the ones in [4] allow each player to privately refine the prior, the revenue benchmarks are still defined with respect to the prior (and the knowledge graphs), without considering the refinements. An important open problem here is to design crowdsourced Bayesian mechanisms whose revenue approximates an even more demanding benchmark — the optimal revenue based on the "aggregated refinement" obtained by combining all players' refinements together.

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] M. Babaioff, N. Immorlica, B. Lucier, and S. M. Weinberg. A simple and approximately optimal mechanism for an additive buyer. In *55rd Annual IEEE Symposium on Foundations of Computer Science (FOCS'14)*, pages 21–30. IEEE, 2014.

[7] C. W. Bach and E. Tsakas. Pairwise epistemic conditions for nash equilibrium. *Games and Economic Behavior*, 85:48–59, 2014.

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

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

[10] 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.

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

[12] 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.

[13] 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.

[14] Y. Cai, N. R. 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)*, pages 926–939, 2016.

[15] G. Carroll and D. Meng. Robust contracting with additive noise. *Journal of Economic Theory*, 166:586–604, 2016.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

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

[22] 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.

[23] R. Cooke. *Experts in uncertainty: opinion and subjective probability in science.* Oxford University Press on Demand, 1991.

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

[25] 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.

[26] 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)*, pages 426–439, 2016.

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

[28] 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.

[29] R. Fagin, J. Y. Halpern, Y. Moses, and M. Vardi. *Reasoning about knowledge*. MIT press, 2004.

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

[31] 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.

[32] 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.

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

[34] 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.

[35] 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.

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

[37] 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.

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

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

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

[41] 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.

[42] N. Miller, P. Resnick, and R. Zeckhauser. Eliciting informative feedback: The peer-prediction method. *Management Science*, 51(9):1359–1373, 2005.

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

[44] D. Prelec. A Bayesian truth serum for subjective data. *Science*, 306(5695):462–466, 2004.

[45] G. Radanovic and B. Faltings. A robust Bayesian truth serum for non-binary signals. In *27th AAAI Conference on Artificial Intelligence (AAAI'13)*, pages 833–839, 2013.

[46] J. Renault and T. Tomala. Repeated proximity games. *International Journal of Game Theory*, 27(4):539–559, 1998.

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

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

[49] 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.

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

[51] A. D. Shaw, J. J. Horton, and D. L. Chen. Designing incentives for inexpert human raters. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, pages 275–284. ACM, 2011.

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

[53] J. Witkowski and D. C. Parkes. Peer prediction without a common prior. In *13th ACM Conference on Electronic Commerce (EC'12)*, pages 964–981, 2012.

[54] 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.

[55] P. Zhang and Y. Chen. Elicitability and knowledge-free elicitation with peer prediction. In *13th International Conference on Autonomous Agents and Multigent Systems (AAMAS'14)*, pages 245–252, 2014.