Better Outcomes from More Rationality*

Jing Chen
CS Dept., Stony Brook University
Stony Brook, NY 11794
jingchen@cs.stonybrook.edu

Silvio Micali
CSAIL, MIT
Cambridge, MA 02139, USA
silvio@csail.mit.edu

Rafael Pass
Dept. of Computer Science
Cornell University
rafael@cs.cornell.edu

Abstract

Mechanism design enables a social planner to obtain a desired outcome by leveraging the players' rationality and their beliefs. It is thus a fundamental, yet unproven, intuition that *the higher the level of rationality of the players, the better the set of obtainable outcomes.*

In this paper we prove this fundamental intuition for players with *possibilistic beliefs*, the traditional model of epistemic game theory. Specifically,

- We define a sequence of *monotonically increasing* revenue benchmarks for single-good auctions, $G^0 \leq G^1 \leq G^2 \leq \cdots$, where each G^i is defined over the players' beliefs and G^0 is the second-highest valuation (i.e., the revenue benchmark achieved by the second-price mechanism).
- We (1) construct a single, interim individually rational, auction mechanism that, *without any clue* about the rationality level of the players, guarantees revenue G^k if all players have rationality levels $\geq k+1$, and (2) prove that no such mechanism can get even close to guarantee revenue G^k when at least two players are at most level-k rational.

Keywords: epistemic game theory, incomplete information, single-good auctions

*The authors intend to publish a one-page abstract at ITCS 2015 if accepted.

1 Introduction

Mechanism design traditionally models beliefs as probability distributions, and the players as expected-utility maximizers. By contrast, epistemic game theory has successfully and meaningfully studied possibilistic (i.e., set-theoretic) beliefs and more nuanced notions of rationality. In this paper we embrace the epistemic model and prove that, in single-good auctions, *"more revenue is obtainable from more rational players."* Let us explain.

Possibilistic (Payoff-type) Beliefs Intuitively, for a player i:

- i's level-0 beliefs consist of his own (payoff) type;
- i's level-1 beliefs consist of the *set* of all type subprofiles of his opponents that he considers possible (although he may be unable to compare their relative likelihood);
- i's level-2 beliefs consist of the set of level-1 belief subprofiles of his opponents that he considers possible;
- and so on.

As usual, beliefs can be wrong[1] *and beliefs of different players may be inconsistent; furthermore, we do not assume the existence of a common prior, or that a designer has information about the players' beliefs.*

Rationality Following Aumann [4], we do not assume that the players are expected utility maximizers, and let them choose actions that are "rational in a minimal sense". Intuitively,

- A player is *(level-1) rational* if he only plays actions that are not *strictly* dominated by some fixed pure action in *every* world he considers possible.[2]
- Recursively, a player is *level-(k+1) rational* if he (a) is rational and (b) believes that all his opponents are level-k rational.

We do not assume that a mechanism (designer) has any information about the players' rationality level.

Intuitive Description of Our Revenue Benchmarks For auctions of a single good, we consider a sequence of demanding revenue benchmarks, $G^0, G^1, \ldots$.

Intuitively, for any non-negative value v,

- $G^0 \geq v$ if and only if there exist at least two players valuing the good at least v.
  (Note that G^0 is the benchmark achieved by the second-price mechanism.)

- $G^1 \geq v$ if and only if there exist at least two players
  believing that there exists a player (whose identity need not be known)
  valuing the good at least v.[3]

- $G^2 \geq v$ if and only if there exist at least two players
  believing that there exists a player (whose identity need not be known)
  believing that there exists a player (whose identity need not be known)
  valuing the good at least v.

- And so on.

[1] That is a player's belief —unlike his *knowledge*— need not include the true state of the world.

[2] Due to this notion of rationality, it is without loss of generality to restrict to possibilistic beliefs. If players had probabilistic beliefs, the support of these beliefs alone determines whether a player is rational.

[3] Note that G^1 is the benchmark achieved by the mechanism of [15].

EXAMPLE 1. Let there be three players, with respective true valuations 0, 50, and 100. Then $G^0 = 50$. □

EXAMPLE 2. Let there be three players, with respective true valuations 0, 0, and 100. Each player believes that the valuation subprofile for his opponents is either $(0, 100)$ or $(100, 0)$.[4] Then $G^0 = 0$ and $G^1 = 100$. □

EXAMPLE 3. Let there be two players, each valuing the good 0, with the following beliefs.
Player 1 believes that player 2

(a) values the good 100 and

(b) believes that player 1 values it 200.

Player 2 believes that player 1

(a') values the good 100 and

(b') believes that player 2 values it 300.

Then $G^0 = 0$, $G^1 = 100$, and $G^2 = 200$. □

It is intuitive (and easily verifiable from the formal definitions of our benchmarks) that

(i) G^0 coincides with the second-highest valuation;

(ii) $G^0 \le G^1 \le \cdots$, and each G^{k+1} can be arbitrarily higher than G^k;

(iii) If the players' beliefs are correct, then each G^k is less than or equal to the highest valuation, and even G^1 can coincide with this valuation;

(iv) If the players' beliefs are wrong, then even G^1 can be arbitrarily higher than the highest valuation.

Our Results We prove that each additional level of rationality enables one to guarantee a stronger revenue benchmark. Intuitively,

Theorem 1 proves the existence of a *single*, interim individually rational mechanism M that, for all k and all $\varepsilon > 0$, guarantees revenue $\ge G^k - \varepsilon$ whenever the players are level-$(k+1)$ rational; and

Theorem 2 proves that, for any k and any $\delta > 0$, no interim individually rational mechanism can guarantee revenue $\ge G^k - \delta$ if at least 2 players are at most level-k rational.

Recall that a mechanism is interim individually rational if each player i, given his true value, has an action guaranteeing him non-negative utility no matter what his opponents might do.

Let us point out that Theorem 1 generalizes to infinitely rational players (see Section 7).

[4] That is, each player believes that one of his opponents values the good for 100 and the other one values it 0, but does not know which of the two has higher valuation.

1.1 Discussion

A Separation Result (and Two Open Problems) Mechanism design enables a social planner to obtain a desired outcome by leveraging the players' rationality and their beliefs. It is thus a fundamental intuition that "the higher the level of rationality of the players, the better the set of obtainable outcomes". Theorems 1 and 2 prove this intuition. Indeed, for the case of revenue in single-good auctions, they separate the power of different player-rationality levels in mechanism design. To the best of our knowledge, no such separation was previously known.[5]

We stress that Theorems 1 and 2 prove the desired separation without figuring out the optimal revenue achievable under each rationality level. Figuring out such revenue remains an open problem.

Another open problem is whether Theorem 2 continues to hold if only a single player is at most level-k rational and all other players are at least level-$(k+1)$ rational.

An Unusual Guarantee The guarantee of Theorem 1 is stronger than "For each k there exists a mechanism M_k guaranteeing revenue $\geq G^k - \varepsilon$ whenever the players are level-$(k+1)$ rational." Indeed, in the latter case, each M_k might know that every player has a rationality level $\geq k+1$.

By contrast, our mechanism M has no information about the players' rationality levels. It *automatically* guarantees revenue $\geq G^k - \varepsilon$ when the rationality level of each player *happens to be* $\geq k+1$. That is, M returns revenue

$\geq G^0 - \varepsilon$ if the players are level-1 rational;
$\geq G^1 - \varepsilon$ if the players are level-2 rational;
$\geq G^2 - \varepsilon$ if the players are level-3 rational;
and so on.

This guarantee is somewhat unusual: typically a mechanism is analyzed under only one specific solution concept, and thus under one specific rationality level.

Leveraging Higher-Level Beliefs Higher-level beliefs routinely affect people's strategic choices. In the stock market traders may buy some stocks at prices higher than they value them only because they believe that someone else (whose identity they may not know) will later on buy those stocks at even higher prices. It is thus important for mechanism design to develop conceptual frameworks and techniques enabling a social planner to use the players' higher-level beliefs in order to achieve more goals. As we shall explain in our next section, *robust mechanisms* [8] consider higher-level beliefs, but not for broadening the set of implementable outcomes. By contrast, mechanism M of Theorem 1 uses higher-level beliefs in order to increase the revenue obtainable in single-good auctions.

Epistemic vs. Bayesian Frameworks Our mechanism M is also applicable in a Bayesian framework, by simply forming the players' possibilistic beliefs as the support of their probabilistic beliefs.[6] In a Bayesian framework, however, the players are assumed to have very structured information about each other. In particular, the level-1 beliefs of a player specify not only which

[5]Indeed, while epistemic game theory has always dealt with *very nuanced* notions of rationality, all mechanisms designed so far envisaged very coarse rationality levels: namely, either 1 or infinite. The only exception was the mechanism of [15], which considered players with rationality level 2.

[6]In Bayesian settings it has been widely assumed that the support of a player's probabilistic beliefs coincides with the whole type space of all players —the "full-support" assumption. Under this assumption the support of a player's probabilistic beliefs do not contain information about others, since he believes that "everything is possible". But the full-support assumption is without loss of generality only when there is a common prior, in which case (a) any state not in the support of the prior can simply be removed from the state space and (b) it is common knowledge that this has been done. If, as in this paper, no common prior is assumed, it is unclear why different players' probabilistic beliefs must have the same support. When the supports are different, they can be very informative.

type subprofiles are possible for his opponents in his mind, but also the exact relative likelihood of any pair of such subprofiles. Thus, a properly chosen Bayesian mechanism should be able to utilize this richer information better than ours, if relying on a stronger notion of rationality (e.g., expected-utility maximization).[7] The advantage of mechanism M is in non-Bayesian settings of incomplete information, when a player is unable to compare the relative likelihood of his opponents' type subprofiles. In such settings our mechanism successfully elicits the players' beliefs whether or not they are consistent with each other, and whether or not they are correct.

Each approach and each mechanism indeed has its own range of applicability.

2 Related Work

Ever since Harsanyi [25], the players' beliefs in settings of incomplete information traditionally use probabilistic representations (see Mertens and Zamir [28], Brandenburger and Dekel [14], and the survey by Siniscalchi [30].)

Beliefs that are not probabilistic and players that do not maximize expected utilities have been considered by Ellsberg [20]. He considers beliefs with ambiguity, but in decision theory. Thus his work does not apply to higher-level beliefs or multi-player games. Higher-level beliefs with ambiguity in multi-player games have been studied by Ahn [1]. His work, however, is not concerned with implementation, and relies on several common knowledge assumptions about the internal consistency of the players' beliefs. Bodoh-Creed [12] characterizes revenue-maximizing single-good auction mechanisms with ambiguity-averse players, but without considering higher-level beliefs, and using a model quite different from ours.[8] For more works on ambiguous beliefs, see Bewley [11] and the survey by Gilboa and Marinacci [23].

As we shall see in a moment, our belief model is a set-theoretic version of Harsanyi's type structures. Set-theoretic information has also been studied by Aumann [3], but assuming that a player's information about the "true state of the world" is always correct. Independently, set-theoretic models of beliefs have been considered, in modal logic, by Kripke [27] (see [22] for a well written exposition).

Robust mechanism design, as initiated by Bergemann and Morris [8], is close in spirit to our work, but studies questions different from ours. In particular, it provides additional justification for implementation in dominant strategies. Although defining social choice correspondences over the players' payoff types only (rather than their arbitrary higher-level beliefs), Bergemann and Morris [9] explicitly point out that such restricted social choice correspondences cannot represent revenue maximizing allocations.

Chen and Micali [15] have considered arbitrary (possibly correlated) valuations in single-good auctions when the players' beliefs are possibilistic. However, their work uses only the players' first two levels of beliefs. Although our mechanism can be viewed as a generalization of theirs, our and their respective analysis are very different. Indeed, we analyze our mechanism using standard epistemic solution concepts with respect to a very weak notion of rationality, whereas [15] introduced a new solution concept which assumes mutual belief of rationality with respect to the players being expected-utility maximizers. In fact, it is easy to see that our notion of level-2 rational implementation (a special case of our notion) implies their notion of conservative strict implementation, but

[7] As mentioned, if we stick to our weak notion of Aumann rationality [4], only the support of the beliefs matters and probabilistic beliefs collapse down to possibilistic ones.

[8] In his model, the players have preferences of the Maximin Expected Utility form, the designer has a prior distribution over the players' valuations, the players' beliefs are always correct (i.e., they all consider the designer's prior plausible), actions coincide with valuations, and the solution concepts used are dominant strategy and Bayesian-Nash equilibrium.

not vice versa.

Finally, it is also easy to see that our notion of level-1 rational implementation implies implementation in undominated strategies [26], but not vice versa.

3 Our Epistemic Model

Our model is directly presented for single-good auctions, although it generalizes simply to other strategic settings.

An auction is decomposed into two parts: a *context*, describing the set of possible outcomes and the players (including their valuations and their beliefs), and a *mechanism*, describing the actions available to the players and the process leading from actions to outcomes.

We focus on contexts with finitely many types and on deterministic normal-form mechanisms assigning finitely many (pure) actions to each player.

Contexts A context C consists of four components, $C = (n, V, \mathcal{T}, \tau)$, where

- n is a positive integer, *the number of players,* and $[n] \triangleq \{1, \dots, n\}$ is *the set of players.*
- V is a positive integer, *the valuation bound.*
- $\mathcal{T}$, the *type space*, is a tuple of profiles $\mathcal{T} = (T, \Theta, \nu, B)$ where for each player i,
  - T_i is a finite set, the set of i's possible *types*;
  - $\Theta_i = \{0, 1, \dots, V\}$ is the set of i's possible *valuations*;
  - $\nu_i : T_i \to \Theta_i$ is i's *valuation function*; and
  - $B_i : T_i \to 2^{T_{-i}}$ is i's *belief correspondence.*
- τ, the *true type profile*, is such that $\tau_i \in T_i$ for all i.

Note that $\mathcal{T}$ is a possibilistic version of Harsanyi's type structure [25]. As usual, in a context $C = (n, V, \mathcal{T}, \tau)$ each player i privately knows his own true type τ_i and his beliefs. Player i's beliefs are *correct* if $\tau_{-i} \in B_i(\tau_i)$. The profile of *true valuations* is $\theta \triangleq (\nu_i(\tau_i))_{i\in[n]}$.

An outcome is a pair (w, P), where $w \in \{0, 1, \dots, n\}$ is the *winner* and $P \in \mathbb{R}^n$ is the *price profile.* If $w > 0$ then player w gets the good, otherwise the good is unallocated. If $P_i \geq 0$ then player i pays P_i to the seller, otherwise i receives $-P_i$ from the seller. Each player i's *utility function* u_i is defined as follows: for each valuation $v \in \Theta_i$ and each outcome (w, P), $u_i(v, (w, P)) = v - P_i$ if $w = i$, and $= -P_i$ otherwise. i's *utility* for an outcome (w, P) is $u_i(\theta_i, (w, P))$, and sometimes written as $u_i(w, P)$. The *revenue* of outcome (w, P), denoted by $rev(w, P)$, is $\sum_i P_i$.

The set of all contexts with n players and valuation bound V is denoted by $\mathscr{C}_{n,V}$.

Mechanisms An auction mechanism M for $\mathscr{C}_{n,V}$ specifies

- The set $A \triangleq A_1 \times \cdots \times A_n$, where each A_i is i's *set of actions.* We set $A_{-i} \triangleq \times_{j\neq i} A_j$.
- An outcome function, typically denoted by M itself, mapping A to outcomes.

For each context $C \in \mathscr{C}_{n,V}$, we refer to the pair (C, M) as an *auction.*

In an auction, when the mechanism M under consideration is clear, for any player i, valuation v, and action profile a, we may simply use $u_i(v, a)$ to denote $u_i(v, M(a))$, and $u_i(a)$ to denote $u_i(M(a))$.

A mechanism is *interim individually rational (IIR)* if, for every context $C = (n, V, \mathcal{T}, \tau)$ and every player i, there exists some action $a_i \in A_i$ such that for every $a_{-i} \in A_{-i}$,

$$u_i(a) \geq 0.$$

Rationality In a normal-form game with possibilistic beliefs, the notion of (higher-level) rationality of our introduction corresponds to a particular iterative elimination procedure of players' actions; and we demonstrate this characterization in a companion paper [16]. Namely, for every rationality level k, the k-round elimination procedure yields the actions compatible with the players being level-k rational, as follows.

Let $\Gamma = ((n, V, \mathcal{T}, \tau), M)$ be a single-good auction, where $\mathcal{T} = (T, \Theta, \nu, B)$. For each player i, each type $t_i \in T_i$ and each $k \geq 0$, we inductively define $RAT_i^k(t_i)$, the *set of level-k rationalizable actions for* t_i, in the following manner:

- $RAT_i^0(t_i) = A_i$.
- For each $k \geq 1$, $RAT_i^k(t_i)$ is the set of actions $a_i \in RAT_i^{k-1}(t_i)$ for which there does not exist an alternative action $a_i' \in A_i$ such that $\forall t_{-i} \in B_i(t_i)$ and $\forall a_{-i} \in RAT_{-i}^{k-1}(t_{-i})$,

$$u_i(\nu_i(t_i)), (a_i', a_{-i})) > u_i(\nu_i(t_i), (a_i, a_{-i}))$$

  where $RAT_{-i}^k(t_{-i}) = \times_{j \neq i} RAT_j^k(t_j)$.

The set of level-k rationalizable action profiles for auction Γ is $RAT^k(\tau) \triangleq \times_i RAT_i^k(\tau_i)$.

Epistemic Implementation An (epistemic) revenue benchmark b is a function mapping contexts to reals.

Definition 1. *A mechanism M* level-k rationally implements *a revenue benchmark b for $\mathscr{C}_{n,V}$ if, for every context $C \in \mathscr{C}_{n,V}$ and every profile a of level-k rationalizable actions in the auction (C, M),*

$$rev(M(a)) \geq b(C).$$

Notice that our notion of implementation does not require the players have the same level of rationality. Since $RAT^{k'}(\tau) \subseteq RAT^k(\tau)$ for any $k' \geq k$, if a mechanism level-k rationally implements b, then it guarantees b as long as all players have rationality levels $\geq k$.

Furthermore, our notion of implementation does *not* depend on common belief of rationality (a very strong assumption); does *not* require any consistency about the beliefs of different players; and is by definition *"closed under Cartesian product."*[9]

Finally, let us stress that in our notion the mechanism knows only the number of players and the valuation bound. (One may consider weaker notions where the mechanism is assumed to know —say— the entire underlying type space, but not the players' true types. Of course more revenue benchmarks might be implementable under such weaker notions.)

4 Our Epistemic Benchmarks

Below we recursively define the epistemic revenue benchmarks G^k for single-good auctions, based on the players' level-k beliefs. Each G^k is a function mapping a context $C = (n, V, \mathcal{T}, \tau)$ to a real number. For simplicity we let $\max\{v\} \triangleq \max\{v_1, \ldots, v_n\}$ for every profile $v \in \mathbb{R}^n$.

[9] For a given solution concept S this means that S is of the form $S_1 \times \cdots \times S_n$, where each S_i is a subset of i's actions. This property is important from an epistemic perspective, because it overcomes the "epistemic criticism" of the Nash equilibrium concept, see [6, 5, 2]. It is also important from an implementation perspective. In particular, implementation at all Nash equilibria is not closed under Cartesian product, and thus mismatches in the players' beliefs (about each other's equilibrium actions) may easily yield undesired outcomes.

Definition 2. *Let $C = (n, V, \mathcal{T}, \tau)$ be a context where $\mathcal{T} = (T, \Theta, \nu, B)$. For each player i and each integer $k \geq 0$, the function g_i^k is defined as follows: $\forall\ t_i \in T_i$,*

$$g_i^0(t_i) = \nu_i(t_i) \quad \text{and} \quad g_i^k(t_i) = \min_{t'_{-i} \in B_i(t_i)} \max\{(g_i^{k-1}(t_i), g_{-i}^{k-1}(t'_{-i}))\}\ \forall k \geq 1.$$

We refer to $g_i^k(t_i)$ as the level-k guaranteed value of i with type t_i.

The level-k revenue benchmark G^k *maps C to the second highest value in $\{g_i^k(\tau_i)\}_{i\in[n]}$.*

For any $\varepsilon > 0$, $G^k - \varepsilon$ is the revenue benchmark mapping every context C to $G^k(C) - \varepsilon$.

Note that, if $g_i^k(t_i) \geq c$, then player i with type t_i believes that there always exists some player $j^{(1)}$ —possibly unknown to i— who believes that there always exists a player $j^{(2)}$ ... who believes that there always exists some player $j^{(k)}$ whose true valuation is at least c.

Example Let us illustrate the above definition, in a simple case involving three players, with the aid of three figures.

Figure 0 shows a context C with its level-0 guaranteed values. Starting with a "world" in which the true-type profile of the players is $\tau = (t_{11}, t_{21}, t_{31})$, the figure shows all the worlds that are possible according to the higher-level beliefs of our three players.

For each type t_{ik} of player i, we explicitly show the corresponding valuation $\nu_i(t_{ik})$ under it. For conciseness, if a type of a player appears multiple times, we show its corresponding valuation only once.

If in a world ω a player i believes that a world ω' is possible, then we draw an arrow, with label i, from ω to ω'. More precisely, there is an arrow labeled by i from a type (sub)profile t to another type subprofile t' if and only if $t' \in B_i(t_i)$.

From Figure 0 it is easy to see that $G^0(C) = 3$: the second highest among 3, 1, and 5.

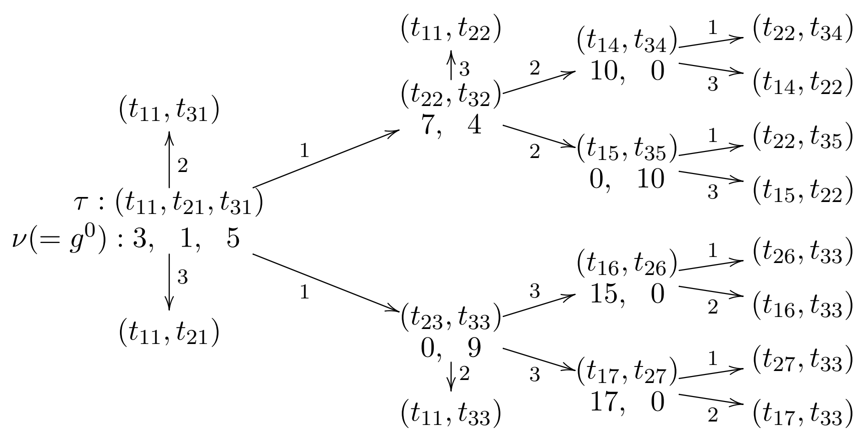

Figure 0: The context C and its level-0 guaranteed values.

From the level-0 guaranteed values of Figure 0, it is easy to compute the level-1 guaranteed values of the worlds in C. In particular,

$$g_1^1(t_{11}) = \min\{\max\{g_1^0(t_{11}), g_2^0(t_{22}), g_3^0(t_{32})\}, \max\{g_1^0(t_{11}), g_2^0(t_{23}), g_3^0(t_{33})\}\} = \min\{7, 9\} = 7.$$

All level-1 guaranteed values are shown in Figure 1. From this figure it is apparent that $G^1(C) = 5$. Indeed, 5 is the second highest value among 7, 5, and 5.

Finally, from the level-1 guaranteed values of Figure 1, it is easy to compute the level-2 guaranteed values in C, shown in Figure 2. From this figure it is apparent that $G^2(C) = 7$.

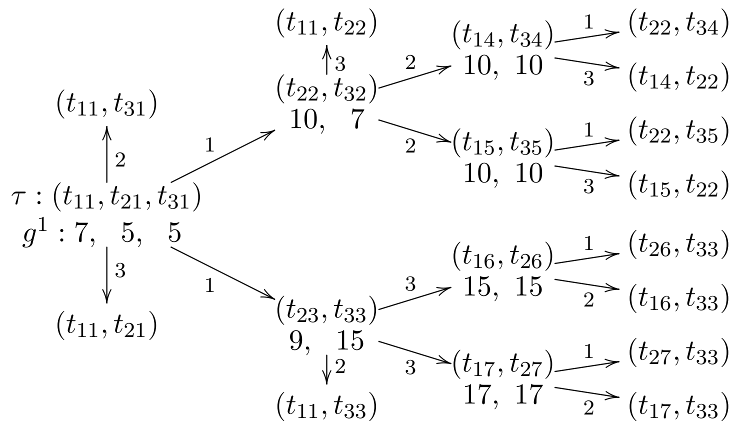

Figure 1: The level-1 guaranteed values of C.

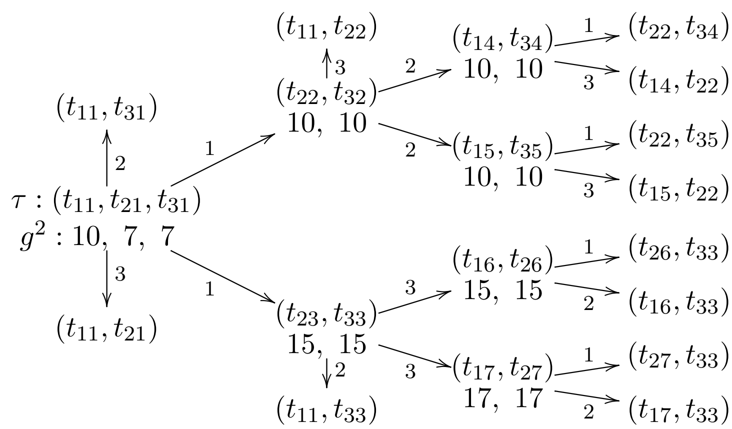

Figure 2: The level-2 guaranteed values of C.

Remark Note that the values g_i^k's are monotonically non-decreasing in k. Indeed,

$$g_i^k(t_i) = \min_{t'_{-i} \in B_i(t_i)} \max\{(g_i^{k-1}(t_i), g_{-i}^{k-1}(t'_{-i}))\} \geq \min_{t'_{-i} \in B_i(t_i)} g_i^{k-1}(t_i) = g_i^{k-1}(t_i).$$

Thus $G^k(C) \geq G^{k-1}(C)$ for every context C and $k > 0$. $G^0(C)$ is the second highest true valuation. It is easy to see that, for every context C, if the players' beliefs are correct, then for each player i and each $k \geq 0$, we have $g_i^k(\tau_i) \leq \max_j \theta_j$, and thus $G^k(C) \leq \max_j \theta_j$.

5 Our First Theorem

While the players' beliefs may be arbitrarily complex, we now show that they can be successfully leveraged by a normal-form mechanism that asks the players to report very little information. Roughly speaking, our mechanism pays the players to receive information about their beliefs, and then uses such information to set a sophisticated reserve price in an otherwise ordinary second-price auction. The idea of buying information from the players is not new (see, e.g., [17], [19], and [15]). We are not aware, however, of any mechanism where higher-level beliefs are being bought. In some sense, our mechanism pays to hear even the *faintest rumors*.

A bit more precisely, the mechanism elicits the players' beliefs up to some *level bound* K that can be arbitrarily high.[10] That is, if $K = 99$, then our mechanism elicits the players' level-0 up to level-99 beliefs about valuations when they happen to be respectively level-1 up to level-100 rational, but does not elicit the players' level-100 beliefs even if they happen to be level-101 rational or more.

Our mechanism is uniformly constructed on parameters n, V, K, and a constant $\varepsilon > 0$. An action of a player i has three components: his own identity (for convenience only), a *belief-level* $\ell_i \in \{0, 1, \ldots, K\}$, and a *value* $v_i \in \{0, 1, \ldots, V\}$. In the description below, the players act only in Step **1**, and Steps **a** through **c** are just "conceptual steps taken by the mechanism".

The expression "$X := x$" denotes the operation that sets or resets variable X to value x.

Mechanism $M_{n,V,K,\varepsilon}$

1: *Each player i, publicly and simultaneously with the others, announces a triple* $(i, \ell_i, v_i) \in \{i\} \times \{0, 1, \ldots, K\} \times \{0, 1, \ldots, V\}$.

a: *Order the n announced triples according to $v_1, \ldots, v_n$ decreasingly, and break ties according to $\ell_1, \ldots, \ell_n$ increasingly. If there are still ties, then break them according to the players' identities increasingly.*

b: *Let w be the player in the first triple,* $P_w := 2^{nd}v \triangleq \max_{j \neq w} v_j$, *and* $P_i := 0 \ \forall i \neq w$.

c: $\forall i$, $P_i := P_i - \delta_i$, *where* $\delta_i \triangleq \frac{\varepsilon}{2n}\left[1 + \frac{v_i}{1+v_i} - \frac{\ell_i}{(1+\ell_i)(1+V)^2}\right]$.

The final outcome is (w, P). We refer to δ_i as player i's reward.

Note that our mechanism never leaves the good unsold.

Remark Allegedly, if i is level-k rational, then $v_i = g_i^{k-1}(\tau_i)$ and $\ell_i = \min\{\ell : g_i^{\ell}(\tau_i) = g_i^{k-1}(\tau_i)\}$. That is, v_i is the highest value v such that i believes *"there exists some player who believes"* ... $(k-1$ times) some player values the good v, and ℓ_i is the smallest level of beliefs about beliefs needed to attain v_i. Roughly speaking, v_i is the highest "rumored" valuation according to player i's level-$(k-1)$ beliefs, and ℓ_i is the "closeness" of the rumor.

Theorem 1. *For each n, V, K and $\varepsilon > 0$, the mechanism $M_{n,V,K,\varepsilon}$ is IIR and, for each $k \in \{0, 1, \ldots, K\}$, level-$(k+1)$ rationally implements the benchmark $G^k - \varepsilon$ for $\mathscr{C}_{n,V}$.*

Note that $M_{n,V,K,\varepsilon}$ does *not* depend on k and is *not* told what the players' rationality level is. Rather, $M_{n,V,K,\varepsilon}$ *automatically* produces revenue $G^k - \varepsilon$ in every play in which the players *happen to be* level-$(k+1)$ rational. Indeed (1) such players use only level-$(k+1)$ rationalizable strategies and, (2) at each profile a of such strategies (as per Definition 1)

$$rev(M_{n,V,K,\varepsilon}(a)) \geq G^k - \varepsilon.$$

In both our intuitive analysis and our proof we arbitrarily fix n, V, K, ϵ and a context $C = (n, V, \mathcal{T}, \tau)$ with $\mathcal{T} = (T, \Theta, \nu, B)$; and simply denote $M_{n,V,K,\varepsilon}$ by M.

[10] The reliance on K is not crucial —in fact, if we are willing to make the action space infinite, then we do not need K and our mechanism can elicit the players' beliefs up to any level.

5.1 Intuition for Theorem 1

Showing that M is IIR is easy. In fact, for each player i, let $a_i \triangleq (i, 0, \theta_i)$. Then i's utility $u_i(a_i, a'_{-i})$ is always non-negative, no matter which action subprofile a'_{-i} the other players choose.

Let us now sketch the proof of our revenue lowerbound, namely,

$$rev(M(a)) \geq G^k(C) - \varepsilon$$

for every $k \in \{0, 1, \ldots, K\}$ and every action profile $a \in RAT^{k+1}(\tau)$.

Notice that

$$v_i \geq g_i^k(\tau_i) \text{ for all } i \quad \text{implies} \quad 2^{nd}v \geq G^k(C),$$

and that the second inequality immediately implies the desired revenue lowerbound, because each reward δ_i is at most $\frac{\varepsilon}{n}$. Therefore it only remains to show that

$$v_i \geq g_i^k(\tau_i) \text{ for every action } a_i = (i, \ell_i, v_i) \in RAT_i^{k+1}(\tau_i).$$

We proceed by contradiction. Assuming $v_i < g_i^k(\tau_i)$, we derive a contradiction by proving the existence of another action $\hat{a}_i$ such that for each type subprofile $t_{-i} \in B_i(\tau_i)$ and each action subprofile $a'_{-i} \in RAT_{-i}^k(t_{-i})$,

$$u_i(a_i, a'_{-i}) < u_i(\hat{a}_i, a'_{-i}).$$

Set $\hat{a}_i = (i, \hat{v}_i, \hat{\ell}_i)$, where $\hat{v}_i = g_i^k(\tau_i)$ and $\hat{\ell}_i = \min\{\ell : g_i^\ell(\tau_i) = g_i^k(\tau_i)\}$, and refer to $\hat{a}_i$ as the *alleged action.*

To begin with, because $\hat{v}_i > v_i$ by construction, no matter what the other players do, using $\hat{a}_i$ gives player i a higher reward than using a_i. But getting a higher reward is not enough to prove the desired inequality. In particular, when $g_i^k(\tau_i) > g_i^0(\tau_i)$, the following may occur.

> *"Bad Case":* Player i does not get the good with a_i, but gets the good and pays a price greater than θ_i with $\hat{a}_i$.

In this case i's utility is positive with a_i, while negative with $\hat{a}_i$. However, we show that the bad case never occurs according to player i's belief. That is, assuming level-$(k+1)$ rationality, we show that

$(*)$ if $g_i^k(\tau_i) > g_i^0(\tau_i)$, then player i believes that he never gets the good by using $\hat{a}_i$.

We derive $(*)$ by proving, by induction, the following two properties: for each player j, each type t_j, and each level-k rationalizable action $a_j = (j, \ell_j, v_j)$,

1. $v_j \geq g_j^{k-1}(t_j)$, and
2. if $v_j = g_j^{k-1}(t_j)$, then $\ell_j \leq \min\{\ell : g_j^\ell(t_j) = g_j^{k-1}(t_j)\}$.

We omit sketching the proofs of these properties, but explain why they imply $(*)$.

By the definition of $g_i^k(\tau_i)$, for any type profile $t = (\tau_i, t_{-i})$ with $t_{-i} \in B_i(\tau_i)$, there exists some player j whose level-$(k-1)$ guaranteed value $g_j^{k-1}(t_j)$ is at least $g_i^k(\tau_i)$. Since i believes that such a player j uses level-k rationalizable actions, by Property 1 he also believes that $v_j \geq g_j^{k-1}(t_j)$. We now distinguish two cases.

If $v_j > g_i^k(\tau_i) = \hat{v}_i$, then of course $j \neq i$, and player i cannot get the good by using $\hat{a}_i$. Thus $(*)$ trivially holds. What if $v_j = g_i^k(\tau_i)$?

In this case, because $v_j \geq g_j^{k-1}(t_j) \geq g_i^k(\tau_i)$, we have $v_j = g_j^{k-1}(t_j)$ as well. According to Property 2, player j, who uses level-k rationalizable actions in i's belief, announces $\ell_j \leq \min\{\ell : g_j^\ell(t_j) = g_j^{k-1}(t_j)\}$. Because $g_i^k(\tau_i) > g_i^0(\tau_i)$, it can be proved that ℓ_j is at most $\hat{\ell}_i - 1$, that is, $\ell_j < \hat{\ell}_i$. Given how the players' announced triples are ordered, j's triple is ordered before i's. Thus i cannot get the good and $(*)$ holds.

To summarize, if player i believes that his opponents are going to use level-k rationalizable actions, then he also believes that it is "safe" for him to use his alleged action, which gives him the biggest reward without any risk of being over-charged. Thus bidding any value strictly less than $g_i^k(\tau_i)$ is interim strictly dominated by the alleged action, and cannot be level-$(k+1)$ rationalizable. This concludes our intuitive analysis.

5.2 Proof of Theorem 1

We break our proof into simpler claims.

Claim 1. *M is IIR.*

Proof. Arbitrarily fix $i \in [n]$ and $a'_{-i} \in A_{-i}$, and let $a_i = (i, 0, \theta_i)$. We need to prove

$$u_i(a_i, a'_{-i}) \geq 0. \tag{1}$$

In the play of (a_i, a'_{-i}), if $w \neq i$, then we have $P_i = -\delta_i$, and thus $u_i(a_i, a'_{-i}) = -P_i = \delta_i > 0$.

If $w = i$, then we have $\theta_i \geq 2^{nd}v$ and $P_i = 2^{nd}v - \delta_i$. Thus

$$u_i(a_i, a'_{-i}) = \theta_i - P_i = \theta_i - 2^{nd}v + \delta_i \geq \delta_i > 0.$$

Therefore Equation 1 holds, and so does Claim 1. □

To prove our revenue lowerbound, we make use of the following relations. For any two pairs of non-negative integers (ℓ, v) and (ℓ', v'), we write

$$(\ell, v) \succ (\ell', v')$$

if $v > v'$ or ($v = v'$ and $\ell < \ell'$). We write $(\ell, v) \succeq (\ell', v')$ if $(\ell, v) \succ (\ell', v')$ or $(\ell, v) = (\ell', v')$.

Claim 2. *Let δ_i and δ'_i respectively be the rewards that player i gets in Step* **c** *according to the action profiles (a_i, a_{-i}) and (a'_i, a_{-i}), where $a_i = (i, \ell_i, v_i)$ and $a'_i = (i, \ell'_i, v'_i)$. Then,*

$$(\ell_i, v_i) \succ (\ell'_i, v'_i) \text{ implies } \delta_i > \delta'_i.$$

Proof. By definition, $(\ell_i, v_i) \succ (\ell'_i, v'_i)$ means that either $v_i > v'_i$, or $v_i = v'_i$ and $\ell_i < \ell'_i$.

If $v_i > v'_i$, then we have

$$\begin{aligned}
\delta_i - \delta'_i &= \frac{\varepsilon}{2n}\left[1 + \frac{v_i}{1+v_i} - \frac{\ell_i}{(1+\ell_i)(1+V)^2}\right] - \frac{\varepsilon}{2n}\left[1 + \frac{v'_i}{1+v'_i} - \frac{\ell'_i}{(1+\ell'_i)(1+V)^2}\right] \\
&= \frac{\varepsilon}{2n}\left[\frac{v_i - v'_i}{(1+v_i)(1+v'_i)} - \frac{\ell_i - \ell'_i}{(1+\ell_i)(1+\ell'_i)(1+V)^2}\right] \\
&> \frac{\varepsilon}{2n}\left[\frac{1}{(1+V)^2} - \frac{\ell_i - \ell'_i}{(1+\ell_i)(1+\ell'_i)(1+V)^2}\right] > \frac{\varepsilon}{2n}\left[\frac{1}{(1+V)^2} - \frac{1}{(1+V)^2}\right] = 0,
\end{aligned}$$

where the first inequality holds because $v'_i < v_i \leq V$, and the second because $\frac{\ell_i - \ell'_i}{(1+\ell_i)(1+\ell'_i)} \leq \frac{\ell_i}{1+\ell_i} < 1$. Thus $\delta_i > \delta'_i$ as desired.

If $v_i = v'_i$ and $\ell_i < \ell'_i$, then we have

$$\delta_i - \delta'_i = \frac{\varepsilon}{2n} \cdot \frac{\ell'_i - \ell_i}{(1+\ell_i)(1+\ell'_i)(1+V)^2} > 0.$$

Thus again $\delta_i > \delta'_i$.

Therefore Claim 2 holds. □

Let us now prove that a player i never "underbids his beliefs".

Claim 3. $\forall\ k \in \{1,\dots,K+1\}$ *and* $\forall a_i = (i,\ell_i,v_i) \in RAT_i^k(\tau_i)$,

$$(\ell_i, v_i) \succeq (\min\{\ell : g_i^\ell(\tau_i) = g_i^{k-1}(\tau_i)\}, g_i^{k-1}(\tau_i)).$$

Proof. We prove Claim 3 by induction on k. Because the analyses for the Base Case ($k=1$) and the Inductive Step ($k>1$) are almost the same, below we focus on the Inductive Step, and point out the differences with the Base Case when needed.

Assume that Claim 3 holds for all $k' < k$. To prove it for k we proceed by contradiction. Letting $\hat{\ell}_i = \min\{\ell : g_i^\ell(\tau_i) = g_i^{k-1}(\tau_i)\}$ and assuming $(\hat{\ell}_i, g_i^{k-1}(\tau_i)) \succ (\ell_i, v_i)$, we shall prove that there is another action $\hat{a}_i$ such that, arbitrarily fixing $t_{-i} \in B_i(\tau_i)$ and $a'_{-i} \in RAT_{-i}^{k-1}(t_{-i})$, we have

$$u_i(\theta_i, (\hat{a}_i, a'_{-i})) > u_i(\theta_i, (a_i, a'_{-i})), \tag{2}$$

contradicting the fact $a_i \in RAT_i^k(\tau_i)$. Let $\hat{v}_i = g_i^{k-1}(\tau_i)$ and set

$$\hat{a}_i \triangleq (i, \hat{\ell}_i, \hat{v}_i).$$

To prove Equation 2, let $\hat{\delta}_i$ and δ_i respectively be the rewards that player i gets in Step **c** in the plays of $(\hat{a}_i, a'_{-i})$ and (a_i, a'_{-i}). Because $(\hat{\ell}_i, \hat{v}_i) \succ (\ell_i, v_i)$, by Claim 2 we have

$$\hat{\delta}_i > \delta_i.$$

Let $(\hat{w}, \hat{P})$ and (w, P) respectively be the outcomes of the two plays, and denote a'_j by (j, ℓ'_j, v'_j) for each $j \neq i$. We distinguish two cases.

Case 1. $\hat{\ell}_i = 0$.

This case applies to both the Base Case ($k=1$) and the Induction Step ($k>1$). In this case we have $\hat{v}_i = g_i^{k-1}(\tau_i) = g_i^0(\tau_i) = \theta_i$, and we further distinguish three subcases.

Subcase 1.1. $w = i$.

In this subcase, we have $\hat{w} = i$ as well, since according to M the triple $(i, \hat{\ell}_i, \hat{v}_i)$ is ordered before (i, ℓ_i, v_i). Therefore $P_i = \max_{j\neq i} v'_j - \delta_i$ and $\hat{P}_i = \max_{j \neq i} v'_j - \hat{\delta}_i$. Accordingly,

$$\begin{aligned} u_i(\theta_i, (\hat{a}_i, a'_{-i})) &= \theta_i - \hat{P}_i = \theta_i - \max_{j\neq i} v'_j + \hat{\delta}_i > \theta_i - \max_{j \neq i} v'_j + \delta_i \\ &= \theta_i - P_i = u_i(\theta_i, (a_i, a'_{-i})), \end{aligned}$$

where the inequality holds because $\hat{\delta}_i > \delta_i$. Thus Equation 2 holds.

Subcase 1.2. $w \neq i$ and $\hat{w} = i$.

In this subcase, $\hat{v}_i \geq \max_{j\neq i} v'_j$, $P_i = -\delta_i$, and $\hat{P}_i = \max_{j\neq i} v'_j - \hat{\delta}_i$. Accordingly,

$$\begin{aligned} u_i(\theta_i, (\hat{a}_i, a'_{-i})) &= \theta_i - \hat{P}_i = \theta_i - \max_{j\neq i} v'_j + \hat{\delta}_i = \hat{v}_i - \max_{j\neq i} v'_j + \hat{\delta}_i \geq \hat{\delta}_i \\ &> \delta_i = -P_i = u_i(\theta_i, (a_i, a'_{-i})), \end{aligned}$$

Thus Equation 2 holds.

Subcase 1.3. $w \neq i$ and $\hat{w} \neq i$.

In this subcase, $P_i = -\delta_i$ and $\hat{P}_i = -\hat{\delta}_i$. Accordingly,

$$u_i(\theta_i, (\hat{a}_i, a'_{-i})) = -\hat{P}_i = \hat{\delta}_i > \delta_i = -P_i = u_i(\theta_i, (a_i, a'_{-i})),$$

and again Equation 2 holds.

Case 2. $\hat{\ell}_i \geq 1$.

This case applies to the Induction Step only. (In the Base Case we have $\hat{\ell}_i = 0$.)

In this case, we shall prove that $\hat{w} \neq i$. To do so, first note that, by the definition of $\hat{\ell}_i$,

$$g_i^{\hat{\ell}_i - 1}(\tau_i) < g_i^{\hat{\ell}_i}(\tau_i). \tag{3}$$

Because $t_{-i} \in B_i(\tau_i)$, we have

$$g_i^{\hat{\ell}_i}(\tau_i) = \min_{t'_{-i} \in B_i(\tau_i)} \max\left\{\left(g_i^{\hat{\ell}_i - 1}(\tau_i), g_{-i}^{\hat{\ell}_i - 1}(t'_{-i})\right)\right\} \leq \max\left\{\left(g_i^{\hat{\ell}_i - 1}(\tau_i), g_{-i}^{\hat{\ell}_i - 1}(t_{-i})\right)\right\}. \tag{4}$$

Combining Equations 3 and 4, we have

$$g_i^{\hat{\ell}_i - 1}(\tau_i) < \max\left\{\left(g_i^{\hat{\ell}_i - 1}(\tau_i), g_{-i}^{\hat{\ell}_i - 1}(t_{-i})\right)\right\}.$$

Letting $t = (\tau_i, t_{-i})$ and $j = \text{argmax}_{r \in [n]} g_r^{\hat{\ell}_i - 1}(t_r)$ with ties broken lexicographically, we have

$$g_j^{\hat{\ell}_i - 1}(t_j) = \max\left\{\left(g_i^{\hat{\ell}_i - 1}(\tau_i), g_{-i}^{\hat{\ell}_i - 1}(t_{-i})\right)\right\}.$$

Accordingly,

$$j \neq i \quad \text{and} \quad g_j^{\hat{\ell}_i - 1}(t_j) \geq g_i^{\hat{\ell}_i}(\tau_i),$$

and thus

$$(\hat{\ell}_i - 1, g_j^{\hat{\ell}_i - 1}(t_j)) \succ (\hat{\ell}_i, g_i^{\hat{\ell}_i}(\tau_i)). \tag{5}$$

Because $\hat{\ell}_i \leq k - 1$ and $a'_j \in RAT_j^{k-1}(t_j)$, we have $a'_j \in RAT_j^{\hat{\ell}_i}(t_j)$. Thus by the inductive hypothesis[11] we have

$$(\ell'_j, v'_j) \succeq (\min\{\ell : g_j^{\ell}(t_j) = g_j^{\hat{\ell}_i - 1}(t_j)\}, g_j^{\hat{\ell}_i - 1}(t_j)) \succeq (\hat{\ell}_i - 1, g_j^{\hat{\ell}_i - 1}(t_j)),$$

which together with Equation 5 implies

$$(\ell'_j, v'_j) \succ (\hat{\ell}_i, g_i^{\hat{\ell}_i}(\tau_i)) = (\hat{\ell}_i, g_i^{k-1}(\tau_i)) = (\hat{\ell}_i, \hat{v}_i). \tag{6}$$

By Equation 6 we have that the triple (j, ℓ'_j, v'_j) is ordered before $(i, \hat{\ell}_i, \hat{v}_i)$ according to M, and thus $\hat{w} \neq i$. Since $(\hat{\ell}_i, \hat{v}_i) \succ (\ell_i, v_i)$, we have $w \neq i$ as well. Therefore $P_i = -\delta_i$ and $\hat{P}_i = -\hat{\delta}_i$, which implies

$$u_i(\theta_i, (\hat{a}_i, a'_{-i})) = -\hat{P}_i = \hat{\delta}_i > \delta_i = -P_i = u_i(\theta_i, (a_i, a'_{-i})).$$

Thus Equation 2 holds.

In sum, Equation 2 holds in all possible cases, contradicting the fact $a_i \in RAT_i^k(\tau_i)$. Therefore Claim 3 holds. □

Following Claim 3, we have that for every action profile $a \in RAT^{k+1}(\tau)$, $2^{nd}v$ is at least the second highest value in the set $\{g_i^k(\tau_i)\}_{i \in [n]}$, which is precisely $G^k(C)$. Because for each player i

$$\delta_i = \frac{\varepsilon}{2n}\left[1 + \frac{v_i}{1 + v_i} - \frac{\ell_i}{(1 + \ell_i)(1 + V)^2}\right] \leq \frac{\varepsilon}{2n} \cdot 2 = \frac{\varepsilon}{n},$$

we have

$$rev(M(a)) = 2^{nd}v - \sum_i \delta_i \geq G^k(C) - \sum_i \delta_i \geq G^k(C) - \sum_i \frac{\varepsilon}{n} = G^k(C) - \varepsilon.$$

This concludes the proof of Theorem 1. ■

[11] Claim 3 is stated with respect to context C and player i. But due to the arbitrary choice of C and i, the claim applies also to context $C' = (n, V, \mathcal{T}, (\tau_{-j}, t_j))$ and player j.

5.3 Discussion of Theorem 1

Right Reasoning and Wrong Beliefs In mechanism M, when a player's beliefs are correct, he always has a "safe" action that is consistent with level-k rationality and gives him positive utility against all other players' actions consistent with level-$(k-1)$ rationality, and our mechanism generates a lot of revenue under such safe actions. However, a player with wrong beliefs may receive negative utility in our mechanism. Indeed, if the players are level-$(k+1)$ rational, their beliefs are wrong, and G^k exceeds the highest valuation, then at least one player has negative utility, because in this case our mechanism generates revenue higher than the highest valuation. Nonetheless, when a player chooses his action, he *believes* that his utility will be non-negative, and thus willingly participates in our mechanism. This situation is not too dissimilar from that of a rational player who willingly enters the stock market, yet might end up losing money if his beliefs are wrong.

(This exploitation by M of the players' wrong beliefs may appear somewhat unfair. However, we should keep in mind that, when constructing a revenue generating mechanism, a designer works for the seller, and would not do his job properly if he leaves some money on the table.)

Let us now point out that M can generate more revenue from players with higher rationality, *even when their beliefs are correct.*

Higher Revenue from Higher Rationality and Correct Beliefs Consider the type space in Figure 3. Both players have correct beliefs at the true type profile τ: player 1 believes that player 2's true type is either t_2 or t_{2_1}, and player 2 believes that player 1's true type is t_1. At τ:

- Player 1 values the good 0 whereas player 2 values it 100.
- Player 2 knows the true state of the world.
- Player 1 believes that either (a) player 2 values the good 100, or (b) player 2 values it 1.
- Player 1 believes that either (a) player 2 values the good 100, or (b) player 2 believes that player 1 values it 2.
- Etc.

Roughly speaking, at τ, if the players are:

- level-1 rational, then M gets bids 0 and 100, and sells the good for 0.
- level-2 rational, then M gets bids 1 and 100, and sells the good for 1.
- level-3 rational, then M gets bids 2 and 100, and sells the good for 2.
- ...
- level-101 rational, then M gets bids 100 and 100, and sells the good for 100.

That is, although the players' values and beliefs do not change, more revenue is extracted by the seller the more rational they are. Moreover, the highest revenue in the type space (i.e., 100) is extracted with a high enough but finite level of rationality (i.e., level-101).

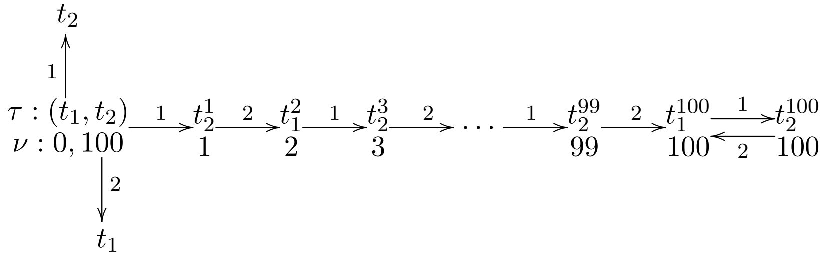

Figure 3: A type space with correct beliefs.

Notice that in the type space of Figure 3, although both players' beliefs are correct, a player may believe that the other may have wrong beliefs.[12] For example, player 1 believes that player 2 may believe that player 1 values the good 2, while player 1 actually values it 0.

Correct Beliefs vs. Common Knowledge of Correct Believes If we considered a type space where at every possible type profile the players have correct beliefs —i.e., in every type profile $t \in \mathcal{T}$, every player i considers t_{-i} possible—, then all our higher-level benchmarks would collapse down to the first level G^1. In this case for our mechanism there is only a gap between the revenue obtained under level-1 rationality and that obtained under level-2 rationality. In fact, level-k rationality, with $k > 2$, yields the same revenue as level-2 rationality. However, in such a type space the players not only have correct beliefs, but actually have *common knowledge of correct beliefs*, which is a much stronger requirement. In particular, standard characterizations of, e.g., rationalizability [13, 31], also do not apply if we restrict to structures where common knowledge of correct beliefs holds [24].

6 Our Second Theorem

Let us now prove that level-$(k+1)$ rationality is necessary to guarantee the benchmark G^k.

Theorem 2. *For every n, V, k, and $c < V$, no IIR mechanism level-k rationally implements $G^k - c$ for $\mathscr{C}_{n,V}$ (even if only two players are level-k rational and all others' rationality levels are arbitrarily higher than k).*

Proof. We first prove the theorem for $n = 2$. Arbitrarily fix $V, k > 0$ (the case where $k = 0$ is degenerated and will be briefly discussed at the end), $c < V$, and an IIR mechanism M. We need to prove the following statement:

$$\text{There exist } C = (2, V, \mathcal{T}, \tau) \in \mathscr{C}_{2,V} \text{ and } a \in RAT^k(\tau) \text{ s.t. } rev(M(a)) < G^k(C) - c. \tag{7}$$

To prove statement 7, we set $\mathcal{T} = (T, \Theta, \nu, B)$ where for each player $i = 1, 2$,

- $T_i = \{t_{i,\ell} : \ell \in \{0, 1, \dots, k\}\}$;
- $\nu_i(t_{i,\ell}) = 0 \ \forall \ell < k$, and $\nu_i(t_{i,k}) = V$; and
- $B_i(t_{i,\ell}) = \{t_{3-i,\ell+1}\} \ \forall \ell < k$, and $B_i(t_{i,k}) = \{t_{3-i,k}\}$.

We set $\tau_i = t_{i,0}$ for each i. The type space $\mathcal{T}$ is illustrated in Figure 4.

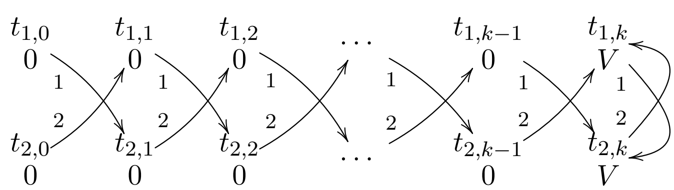

Figure 4: Type space $\mathcal{T}$

Let us now introduce an auxiliary type space $\mathcal{T}' = (T', \Theta, \nu', B')$ where for each player i,

- $T'_i = \{t'_{i,\ell} : \ell \in \{0, 1, \dots, k\}\}$;
- $\nu'_i(t'_{i,\ell}) = 0 \ \ \forall \ell$; and
- $B'_i(t'_{i,\ell}) = \{t'_{3-i,\ell+1}\} \ \forall \ell < k$, and $B'_i(t'_{i,k}) = \{t'_{3-i,k}\}$.

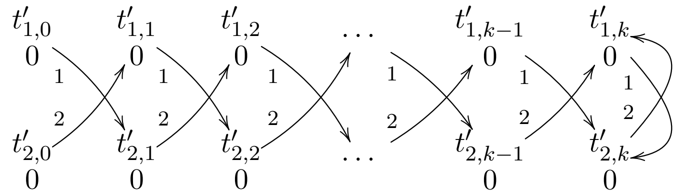

Figure 5: Type space $\mathcal{T}'$

Let $C' = (2, V, \mathcal{T}', \tau')$ where $\tau'_i = t'_{i,0}$ for each i. The type space $\mathcal{T}'$ is illustrated in Figure 5.

In context C, we have $g_i^0(t_{i,k}) = g_i^1(t_{i,k-1}) = \cdots = g_i^{k-1}(t_{i,1}) = g_i^k(t_{i,0}) = V$ for each i. Thus

$$G^k(C) = V, \quad \text{and} \quad G^k(C) - c = V - c > 0.$$

Accordingly, to prove statement 7 it suffices to prove the following two propositions:

$$RAT^k(\tau) = RAT^k(\tau'); \tag{8}$$

and

$$\textit{there exists } a \in RAT^k(\tau') \textit{ such that } rev(M(a)) \leq 0. \tag{9}$$

To prove Equation 8, recall that by definition

$$RAT_i^0(t_{i,\ell}) = RAT_i^0(t'_{i,\ell}) = A_i \text{ for each } i \text{ and each } \ell \leq k,$$

where A_i is the set of actions for player i in M. Because $\nu_i(t_{i,\ell}) = \nu'_i(t'_{i,\ell}) = 0$ for each i and each $\ell < k$, according to our iterated deletion procedure and the construction of $\mathcal{T}$ and $\mathcal{T}'$, by induction we have that for each $\ell' \leq k$,

$$RAT_i^{\ell'}(t_{i,\ell}) = RAT_i^{\ell'}(t'_{i,\ell}) \text{ for each } i \text{ and each } \ell \leq k - \ell'.$$

In particular, for $\ell' = k$ we have $RAT_i^k(t_{i,0}) = RAT_i^k(t'_{i,0})$, that is, $RAT_i^k(\tau_i) = RAT_i^k(\tau'_i)$, for each i. Thus Equation 8 holds.

To prove statement 9, note that $\tau'_i = 0$ for each i. Thus for each action profile a, we have $rev(M(a)) = -u_1(0, a) - u_2(0, a)$. Accordingly, it suffices to prove the following statement:

$$\textit{there exists } a \in RAT^k(\tau') \textit{ such that } u_i(0, a) \geq 0 \textit{ for each } i. \tag{10}$$

To do so, note that M is IIR, which implies that for each player $i = 1, 2$ there exists an action a_i such that

$$u_i(0, (a_i, a'_{3-i})) \geq 0 \ \forall a'_{3-i} \in A_{3-i}.$$

This equation and the definition of $RAT_i^1(\tau'_i)$ together imply that for each i there exists an action $a_i^1 \in RAT_i^1(\tau'_i)$ such that

$$u_i(0, (a_i^1, a'_{3-i})) \geq 0 \ \forall a'_{-i} \in A_{3-i} = RAT_{3-i}^0(t'_{3-i,1}).$$

(Indeed, if $a_i \in RAT_i^1(\tau'_i)$ then $a_i^1 = a_i$, else a_i^1 is the action interim strictly dominating a_i.)

Because $B'_i(\tau'_i) = B'_i(t'_{i,0}) = \{t'_{3-i,1}\}$, by induction we conclude that for each i there exists an action $a_i^k \in RAT_i^k(\tau'_i)$ such that

$$u_i(0, (a_i^k, a'_{3-i})) \geq 0 \ \forall a'_{3-i} \in RAT_{3-i}^{k-1}(t'_{3-i,1}).$$

[12]Indeed, the players' beliefs are correct if they believe that the true world is possible, not that it is the only possible world.

Note that $a^k \in RAT^k(\tau')$. Accordingly, to prove Statement 10 it suffices to show that $a^k_{3-i} \in RAT^{k-1}_{3-i}(t'_{3-i,1})$ for each i, or equivalently,

$$a^k_i \in RAT^{k-1}_i(t'_{i,1}) \ \forall i, \tag{11}$$

because then we have $u_i(0, a^k) \geq 0$ for each i, as desired. To prove Equation 11, again recall that by definition

$$RAT^0_i(t'_{i,\ell}) = RAT^0_i(t'_{i,\ell+1}) = A_i \text{ for each } i \text{ and each } \ell < k.$$

Because the players' valuations are always 0 in $\mathcal{T}'$, we have

$$RAT^1_i(t'_{i,\ell}) = RAT^1_i(t'_{i,\ell+1}) \text{ for each } i \text{ and each } \ell < k - 1.$$

By induction, we finally have

$$RAT^{k-1}_i(t'_{i,0}) = RAT^{k-1}_i(t'_{i,1}) \text{ for each } i.$$

Accordingly, we have $a^k_i \in RAT^k_i(\tau'_i) = RAT^k_i(t'_{i,0}) \subseteq RAT^{k-1}_i(t'_{i,0}) = RAT^{k-1}_i(t'_{i,1})$ for each i. Thus Equation 11 holds, and so does statement 10 and statement 9.

Combining Equation 8 and statement 9, we have that statement 7 holds, and thus Theorem 2 holds for $n = 2$ and $k > 0$.

In the degenerated case where $n = 2$ and $k = 0$, the analysis is very similar. We consider context $C = (2, V, \mathcal{T}, \tau)$ with $\mathcal{T} = (T, \Theta, \nu, B)$, such that for each player i:

$$T_i = \{t_i\}; \quad \nu_i(t_i) = V; \quad \text{and} \quad B_i(t_i) = \{t_{3-i}\}.$$

Also consider the auxiliary context $C' = (2, V, \mathcal{T}', \tau')$ with $\mathcal{T}' = (T', \Theta, \nu', B')$, such that for each player i:

$$T'_i = \{t'_i\}; \quad \nu'_i(t'_i) = 0; \quad \text{and} \quad B'_i(t'_i) = \{t'_{3-i}\}.$$

Because M is IIR, in auction (C', M) there exists an action profile a such that $u_i(0, a) \geq 0$ for each i. But then $rev(M(a)) \leq 0 < V - c = G^0(C) - c$. Because $a \in A = RAT^0(\tau)$, M cannot level-0 rationally[13] implement $G^0 - c$.

In sum, Theorem 2 holds for $n = 2$. For $n > 2$, we construct the desired type spaces (and contexts) by adding dummy players to the type spaces $\mathcal{T}$ and $\mathcal{T}'$ of the 2-player case. The analysis is essentially the same, and thus omitted. ■

7 Variants, Extensions, and Conclusions

Different Reward Functions The total reward given to the players by our mechanism is upper-bounded by an absolute value $\varepsilon > 0$. A similar analysis shows that the mechanism could choose to reward the players with an ε fraction of the price charged to the winner. In this case, the guaranteed revenue would be $(1 - \varepsilon)G^k$ rather than $G^k - \varepsilon$.

[13] Level-0 rationality naturally means that the players are "irrational" and may use any actions.

Infinitely Rational Players Some readers may wonder how much revenue our mechanism generates if the players are infinitely rational —that is, level-k rational for every $k \geq 0$. To answer this question, let $g_i^\infty = \max_k g_i^k$ for each i, and let G^∞ be the second highest of the g_i^∞'s. Since the highest value the players may have is upper bounded by V, each g_i^∞ is finite and can be attained at some finite belief level k_i.[14] Roughly speaking, g_i^∞ is the highest "rumored" valuation according to player i's beliefs and k_i is the "closeness" of the rumor.

To leverage the players' infinitely high rationality levels, *without having any information about such k_i's!*, our mechanism is almost the same as before, except that in Step **1**, each player i announces $\ell_i \in \mathbb{Z}^+$, where $\mathbb{Z}^+$ is the set of non-negative integers. Thus, allegedly, each player i announces (a) $v_i = g_i^\infty$, the highest value v such that i believes *"there exists some player who believes"* ... some player values the good v, and (b) $\ell_i = k_i$, the smallest level of beliefs about beliefs needed to attain v_i. The analysis of the mechanism is almost the same. In particular, our mechanism guarantees the revenue benchmark $G^\infty - \varepsilon$ under common knowledge of rationality.

Conclusions Although studied for generating revenue in single-good auctions, our approach is quite general. In applications where the setting is not Bayesian, it may be important to consider the players' higher-level set-theoretic beliefs. Indeed, attractive social choice correspondences defined over such beliefs may be studied and successfully implemented.

Acknowledgements

The authors thank Shafi Goldwasser, Aviad Heifetz, Andrew Lo, and Ron Rivest for many helpful comments. The third author thanks Joseph Halpern for introducing him to the area of epistemic game theory, and for hours and hours of enlightening discussions about it. The first two authors have been partially supported by ONR Grant No. N00014-09-1-0597. The third author has been partially supported by an Alfred P. Sloan Fellowship, Microsoft New Faculty Fellowship, NSF Award CNS-1217821, NSF CAREER Award CCF-0746990, NSF Award CCF-1214844, AFOSR YIP Award FA9550-10-1-0093, and DARPA and AFRL under contract FA8750-11-2-0211.

References

[1] D. S. Ahn. Hierarchies of ambiguous beliefs. *Journal of Economic Theory*, Vol. 136, pp. 286-301, 2007.

[2] G. B. Asheim, M. Voorneveld, and J. W. Weibull. Epistemically stable action sets. Working paper, 2009.

[3] R. Aumann. Agreeing to Disagree. *Annals of Statistics*, Vol. 4, pp. 1236-1239, 1976.

[4] R. Aumann. Backwards Induction and Common Knowledge of Rationality. *Games and Economic Behavior*, Vol. 8, pp. 6-19, 1995.

[5] R. Aumann and A. Brandenburger. Epistemic Conditions for Nash Equilibrium. *Econometrica*, Vol. 63, No. 5, pp. 1161-1180, 1995.

[6] K. Basu and J.W. Weibull. Action subsets closed under rational behavior. *Economics Letters*, Vol. 36, pp. 141-146, 1991.

[7] P. Battigalli and M. Siniscalchi. Rationalization and incomplete information. *B. E. Journal of Theoretical Economics*, Vol. 3, Iss. 1, 2003.

[14] This is true even when the type space is infinite.

[8] D. Bergemann and S. Morris. Robust mechanism design. *Econometrica*, Vol. 73, No. 6, pp. 1771-1813, 2005.

[9] D. Bergemann and S. Morris. Robust Mechanism Design: An Introduction. In D. Bergemann and S. Morris, *Robust Mechanism Design*, World Scientific Press, 2012.

[10] B. Bernheim. Rationalizable Strategic Behavior. *Econometrica*, Vol. 52, No. 4, pp. 1007-1028, 1984.

[11] T. F. Bewley. Knightian decision theory. Part I. *Decisions in Economics and Finance*, Vol. 25, pp. 79-110, 2002.

[12] A. L. Bodoh-Creed. Ambiguous beliefs and mechanism design. *Games and Economic Behavior*, Vol. 75, pp. 518-537, 2012.

[13] A. Brandenburger and E. Dekel. Rationalizability and correlated equilibria. *Econometrica*. Vol. 55, pp. 1391-1402, 1987.

[14] A. Brandenburger and E. Dekel. Hierarchies of beliefs andcommonknowledge. *Journal of Economic Theory*, Vol. 59, pp. 189-198, 1993.

[15] J. Chen and S. Micali. Mechanism Design with Set-Theoretic Beliefs. *Symposium on Foundations of Computer Science (FOCS)*, pp. 87-96, 2011. An improved version will appear in the *Journal of Economic Theory*.

[16] J. Chen, S. Micali and R. Pass. Possibilistic beliefs and higher-level rationality. *Technical report, MIT-CSAIL-TR-2014-013, 2014*. Available at http://hdl.handle.net/1721.1/87727.

[17] J. Crémer and R. P. Mclean. Optimal Selling Strategies under Uncertainty for a Discriminating Monopolist when Demands are Interdependent. *Econometrica*, Vol. 53, No. 2, pp. 345-361, March 1985.

[18] E. Dekel, D. Fudenberg, S. Morris. Interim correlated rationalizability. *Theoretical Economics*, Vol. 2, pp. 15-40, 2007.

[19] K. Eliaz and R. Spiegler. A Mechanism-Design Approach to Speculative Trade. *Econometrica*, Vol. 75, No. 3, pp. 875-884, 2007.

[20] D. Ellsberg. Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics*, Vol. 75, pp. 643-669, 1961.

[21] J. C. Ely and M. Pęski. Hierarchies of belief and interim rationalizability. *Theoretical Economics*, Vol. 1, pp. 19-65, 2006.

[22] R. Fagin, J. Halpern, Y. Moses, and M. Vardi. *Reasoning About Knowledge*. MIT Press, 2003.

[23] I. Gilboa and M. Marinacci. Ambiguity and the Bayesian paradigm. Working paper, 2011.

[24] J. Halpern and R. Pass. Conservative belief and rationality. *Games and Economic Behavior*, Vol. 80, pp. 186-192, 2013.

[25] J. Harsanyi. Games with Incomplete Information Played by "Bayesian" Players, I-III. *Management Science*, Vol. 14, pp. 159-182, 320-334, 486-502. 1967-1968.

[26] M. Jackson. Implementation in Undominated Strategies: A Look at Bounded Mechanisms. *The Review of Economic Studies*, Vol. 59, pp. 757-775, 1992.

[27] S. Kripke. Semantical analysis of modal logic I: normal modal propositional calculi. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, Vol. 9, pp. 67-96, 1963.

[28] J. Mertens and S. Zamir. Formulation of Bayesian analysis for games with incomplete information. *International Journal of Game Theory*, Vol. 14, pp. 1-29, 1985.

[29] D. Pearce. Rationalizable strategic behavior and the problem of perfection. *Econometrica*, Vol. 52, No. 4, pp. 1029-1050, 1984.

[30] M. Siniscalchi. Epistemic Game Theory: Beliefs and Types. In S. N. Durlauf and L. E. Blume (eds.), *The New Palgrave Dictionary of Economics*, Second Edition. Palgrave Macmillan, 2008.

[31] T. Tan and S. Werlang. The Bayesian foundation of solution concepts of games. *Journal of Economic Theory*, Vol. 45, pp. 370-391, 1988.