Epistemic Implementation: Leveraging Arbitrary Set-Theoretic Belief Hierarchies

Jing Chen
CSAIL, MIT
Cambridge, MA 02139, USA
jingchen@csail.mit.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

August 27, 2012

Abstract

In settings of incomplete information, we model the hierarchy of the players' beliefs about each other's payoff types in a set-theoretic way, and leverage it to generate revenue in single-good auctions. Our mechanisms have no clue about the players' valuations or beliefs. Yet, they work even when a player's beliefs are totally arbitrary and wrong, the beliefs of different players are inconsistent with each other, and the players are not expected-utility maximizers.

For each $k \geq 0$, we define a revenue benchmark G^k over the players' order-k beliefs. Our benchmarks are very demanding: they are monotonically non-decreasing in k; G^0 is the second highest true valuation; and the gap between each pair G^k and G^{k+1} can be arbitrary. In particular, for each $k \geq 1$, G^k can be arbitrarily larger than the highest true valuation when the players' beliefs are wrong, and can coincide with the highest true valuation when all beliefs are correct.

We construct a single, interim individually rational (IIR) mechanism guaranteeing revenue $\geq G^k - \varepsilon$ for all k and all profiles of order-$(k+1)$ rationalizable actions, where the underlying notion of rationality is the very weak one proposed by Aumann [4].

Finally, we separate the revenue achievable from order-k and order-$(k+1)$ rationalizable actions. Indeed we prove that, for any $c > 0$, no IIR mechanism can guarantee revenue $\geq G^k - c$ when the played actions are at most order-k rationalizable.

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

1 Introduction

Weak notions of rationality and set-theoretic beliefs have been long studied in epistemic game theory. However, mechanism design traditionally assumes that the players are expected-utility maximizers, and models their beliefs as probability distributions. In this paper, we take an epistemic game-theoretic approach to design normal-form auctions that generate revenue by leveraging the players' (arbitrary) beliefs about each other's valuations. Intuitively, we adopt the following model of beliefs and rationality:

- A player's order-0 beliefs consist of his own (payoff) type; his order-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); his order-2 beliefs consist of the set of order-1 belief subprofiles of his opponents that he considers possible; and so on.
  We do not require a player's beliefs to be correct, nor the beliefs of different players to be consistent with each other, and we do not assume that a mechanism (designer) has any information about the players' beliefs.
- Our players are not required to be expected-utility maximizers. Following Aumann [4], an action a_i of a player i is *(order-1) rationalizable* if, for every pure action a_i', there exists *some* "state of the world that i considers possible" where a_i performs as well as a_i'. We do not need to assume distributions over states: it suffices to work with "possibilistic" beliefs. An action a_i is *order-$(k+1)$ rationalizable* if it is rationalizable when i believes that all other players use order-k rationalizable actions.

In this model we focus on the problem of generating revenue in single-good auctions, with private values and quasi-linear utilities, via mechanisms that are *interim individually rational* (IIR for short). In such mechanisms, given his own type, a player always has an action guaranteeing him non-negative utility no matter what his opponents might do.

We put forward a sequence of very ambitious revenue benchmarks, $G^0, G^1, \ldots$, where each G^k is defined over the players' order-k beliefs, such that:

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

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

(c) 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;

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

For these benchmarks we prove the following two results:

1. *For every $\varepsilon > 0$, there exists an IIR mechanism M_ε guaranteeing revenue $\geq G^k - \varepsilon$ for every profile of order-$(k+1)$ rationalizable actions*, but

2. *For every $c > 0$, no IIR mechanism can guarantee revenue $\geq G^k - c$ when the players use order-k rationalizable actions.*

Remarks About Our First Result While the players' beliefs may be arbitrarily complex, our mechanism asks them to report very little information. Roughly speaking, M_ε is a second-price auction with a reserve price. The mechanism pays the players to receive information about their beliefs, and then uses such information to set the reserve price. The idea of buying information from the players is not new. (In particular, it is used by the auction mechanism of [14].) We are not aware, however, of any mechanism where higher-order beliefs are being bought. In some sense, our mechanism pays to hear even the *faintest rumors*.

Let us point out that a player may receive negative utility in our mechanism. Indeed, if the players are order-$(k+1)$ 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.)

Finally, let us stress that our first result is stronger than saying that "for every k there exists a mechanism $M_{\varepsilon,k}$ that guarantees revenue $\geq G^k - \varepsilon$." Indeed, we need not know what the rationality order of our players is. Our mechanism *automatically* guarantees revenue $\geq G^0 - \varepsilon$ if the players are order-1 rational; revenue $\geq G^1 - \varepsilon$ if they are order-2 rational; revenue $\geq G^2 - \varepsilon$ if they are order-3 rational; and so on. This guarantee is somewhat unusual, as typically a mechanism is analyzed under a specific solution concept, and thus under a specific rationality order.

Remarks About Our Second Result Notice that prior mechanisms required at most order-2 rationality, or common belief of rationality, but nothing in between. Our second result proves a fundamental intuition: namely, that "each additional rationality order strictly increases implementation power." Indeed, each benchmark G^k separates what is implementable with order-k rationality from what is implementable with order-$(k+1)$ rationality.

Our Approach as an Alternative Road Dominant-strategy and Bayesian mechanisms are the typical approaches to settings of incomplete information, and single-good auctions are no exception. We thus wish to contrast such mechanisms with ours.

The revenue of our mechanism is always at least the second-highest valuation, and sometimes arbitrarily higher. By contrast, the revenue of the second-price mechanism always coincides with the second-highest valuation. The latter mechanism however does not rely on higher-order rationality and does not depend on the players' beliefs.

Our mechanism is also applicable in a Bayesian setting, because the players' set-theoretic beliefs are always defined (in particular, their order-1 beliefs can be taken to consist of the supports of the relevant distributions). In a Bayesian setting, however, the players are assumed to have very structured information about each other. In particular, the order-1 beliefs of a player specify not only which 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 leverage this richer information better than ours.

The real advantage of our mechanism is in non-Bayesian settings, when a player is unable to compare the relative likelihood of his opponents' type subprofiles. In such settings our mechanism successfully leverages the players' beliefs whether or not they are consistent with each other, and whether or not they are correct.

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

2 Related Work

Ever since Harsanyi [21], the players' beliefs in settings of incomplete information traditionally use probabilistic representations (see Mertens and Zamir [24], Brandenburger and Dekel [13], and the survey by Siniscalchi [26].)

Beliefs that are not probabilistic and players who do not maximize expected utilities have been considered by Ellsberg [16]. He considers beliefs with ambiguity, but in decision theory. Thus his work does not apply to higher-order beliefs or multi-player games. Higher-order 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-order beliefs, and using a model quite different from ours.[1] For more works on ambiguous beliefs, see Gilboa and Schmeidler [20], Bewley [11], and the survey by Gilboa and Marinacci [19].

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 [23] (see [18] for a well written exposition).

Rationalizability was defined by Pearce [25] and Bernheim [10] for complete-information settings. We extend it to our setting, via an iterated elimination procedure similar to that proposed by Dekel, Fudenberg, and Morris [15] in a Bayesian setting. For other iterated elimination procedures and corresponding notions of rationalizability in Bayesian settings, see Battigalli and Siniscalchi [7], Ely and Pęski [17], and Weinstein and Yildiz [27].

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-order beliefs), Bergemann and Morris [9] explicitly point out that such restricted social choice correspondences cannot represent revenue maximizing allocations.

Chen and Micali [14] have considered arbitrary (possibly correlated) valuations in single-good auctions when the players' beliefs are possibilistic. However, their work leverages only the players' first two orders 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 [14] 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 order-2 rational implementation implies their notion of conservative strict implementation, but not vice versa.

Finally, it is also easy to see that order-1 rational implementation (a special case of our notion) implies implementation in undominated strategies [22], but not vice versa.

[1] 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.

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 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 $[n] \triangleq \{1, \ldots, n\}$ 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*;
  - $\Theta_i = \{0, 1, \ldots, V\}$ 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 [21]. 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 $w \in \{0, 1, \ldots, n\}$ 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 Essentially, an action is order-k rationalizable if it survives the first k steps of iterated deletion of interim strictly dominated actions. Let us be more precise.

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 define $RAT_i^k(t_i)$, the *set of order-k rationalizable actions for player i with type* t_i, inductively as follows:

- $RAT_i^0(t_i) = A_i$.
- 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 order-k rationalizable action profiles for auction Γ is $RAT^k(\tau) \triangleq \times_i RAT_i^k(\tau_i)$.

Every player is *order-0 rational.* A player is *order-*$(k+1)$ *rational* if he uses order-$(k+1)$ rationalizable actions and believes his opponents to be order-k rational.

Note that our elimination procedure is consistent with Aumann's notion of rationality and higher-order rationality [4]. Also, if we allow an action to be dominated by a mixed action, then our notion can be alternatively defined by eliminating never-best-responses in a way analogous to that of interim correlated rationalizability, as proposed by Dekel, Fudenberg, and Morris [15].

Finally, note that each player can, for every k, compute his order-k rationalizable actions based on his true type τ_i and his beliefs.

Epistemic Implementation An (epistemic) revenue benchmark b is a function mapping contexts to reals. A mechanism M *order-k rationally implements* b for $\mathscr{C}_{n,V}$ if, for every context $C \in \mathscr{C}_{n,V}$ and every profile a of order-k rationalizable actions in the auction (C, M),

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

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."*[2]

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' order-k beliefs. Each G^k is a function mapping a context

[2] 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.

$C = (n, V, \mathcal{T}, \tau)$ to a real number. As a warm-up, we first informally describe G^0, G^1 and G^2.

- Let $g_i^0 = \theta_i$, the true valuation, for each player i. (The interpretation of g_i^0 is that player i "believes" that there exists some player —i.e., himself!— who values the good for at least g_i^0.)
  Then G^0 is defined to be the second highest value among all values g_i^0.
- Let g_i^1 be the highest value c such that player i believes that, no matter what the true type profile may be, there always exists some player j (whose identity need not be known to i) with $g_j^0 \geq c$.
  Then G^1 is defined to be the second highest value among all values g_i^1.
- Let g_i^2 be the highest value c such that player i believes that there always exists some player j (whose identity need not be known to i) with $g_j^1 \geq c$.
  Then G^2 is defined to be the second highest value among all values g_i^2.

Note that G^0 clearly coincides with the second highest true valuation. Also note that, since we allow them to be arbitrary, the players' beliefs can be *totally wrong.* In this case, for $k > 0$, G^k may vastly exceed the highest true valuation. For instance, consider the case of two players, both valuing the good for 10, where player 1 believes that player 2 values the good for at least 200, and player 2 believes that player 1 values it for 300. Then $G^1 = 200$.

We now provide the formal definition.

Definition 1. *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: $\forall\ t_i \in T_i$,*

$$g_i^0(t_i) = \nu_i(t_i) \quad \text{and} \quad g_i^k(t_i) = \min_{t_{-i} \in B_i(t_i)} \max_{j \in [n]} g_j^{k-1}(t_j)\ \forall k \geq 1.$$

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

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

We so name $g_i^k(t_i)$ because, if $g_i^k(t_i) \geq c$ then player i with type t_i believes that there always exists some player $j^{(1)}$ 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.

The g_i^k's are monotonically non-decreasing in k. Indeed, for each player i, integer $k > 0$ and type $t_i \in T_i$, we have

$$g_i^k(t_i) = \min_{t_{-i} \in B_i(t_i)} \max_{j \in [n]} g_j^{k-1}(t_j) \geq \min_{t_{-i} \in B_i(t_i)} g_i^{k-1}(t_i) = g_i^{k-1}(t_i).$$

Thus $G^k(C) \geq G^{k-1}(C)$ for every context C and $k > 0$.

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 $G^k(C) \leq \max_j \theta_j$.

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

5 Our Mechanism

We now construct a mechanism that leverages the players' beliefs up to some *order bound* K that can be arbitrarily high.[3] That is, if $K = 99$, then our mechanism leverages the players'

[3] 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 leverage the players' beliefs up to any order.

order-0 up to order-99 beliefs about valuations when they happen to be respectively order-1 up to order-100 rational, but does not leverage the players' order-100 beliefs even if they happen to be order-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-order* $\ell_i \in \{0, 1, \ldots, K\}$, 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* $(i, \ell_i, v_i) \in \{i\} \times \{0, 1, \ldots, K\} \times \{0, 1, \ldots, V\}$.

(Allegedly, if i is order-k rational, then $\ell_i = \min\{\ell : g_i^\ell(\tau_i) = g_i^{k-1}(\tau_i)\}$ and $v_i = g_i^\ell(\tau_i)$.)

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.

5.1 Analysis of Our Mechanism

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

Both in 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.

5.1.1 Intuitive Analysis

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 order-$(k+1)$ 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 order-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 order-$(k-1)$ 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 order-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 order-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 order-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 order-$(k+1)$ rationalizable. This concludes our intuitive analysis.

5.1.2 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. $\forall\ k \in \{1, \ldots, K+1\}$ *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,

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

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 $\hat{\ell}_i$,

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

Let t be the type profile (τ_i, t_{-i}). Because $t_{-i} \in B_i(\tau_i)$, we have

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

Combining Equations 3 and 4, we have

$$g_i^{\hat{\ell}_i - 1}(\tau_i) < \max_{j'} g_{j'}^{\hat{\ell}_i - 1}(t_{j'}).$$

Therefore, letting $j = \operatorname{argmax}_{j'} g_{j'}^{\hat{\ell}_i - 1}(t_{j'})$ with ties broken lexicographically, we have

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

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[4] 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. ■

[4] 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.

6 Impossibility Results for Epistemic Implementation

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

Theorem 2. *For every n, V, k, and $c < V$, no IIR mechanism order-k rationally implements $G^k - c$ for $\mathscr{C}_{n,V}$.*

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:

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

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

- $T_i = \{t_{i,\ell} : \ell \in \{0, 1, \ldots, k\}\}$;
- $\nu_i(t_{i,\ell}) = 0 \ \forall \ell < k$, 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 1, where a circle represents a type, the number inside a circle represents the corresponding valuation, and the arrows represent the belief correspondences.

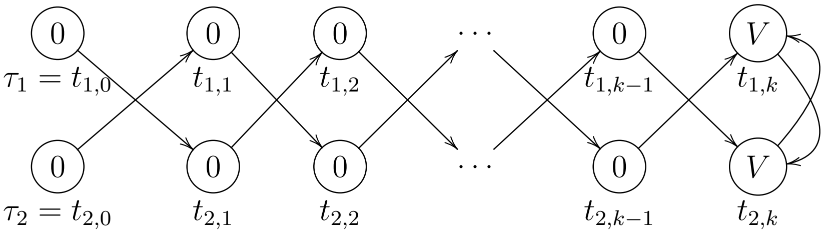

Figure 1: Type space $\mathcal{T}$

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

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

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 2.

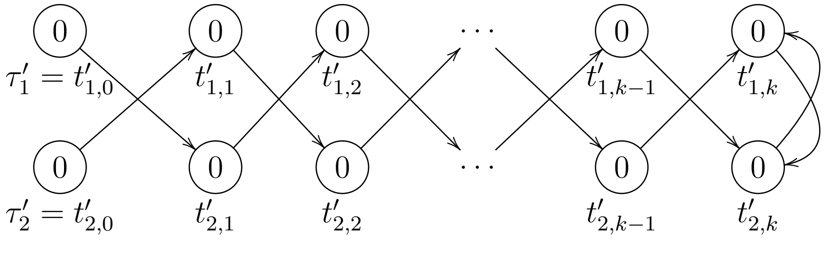

Figure 2: Type space $\mathcal{T}'$

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

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

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

$$a_i^k \in RAT_i^{k-1}(t'_{i,1}) \ \forall i, \tag{11}$$

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

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

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

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

By induction, we finally have

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

Accordingly, we have $a_i^k \in RAT_i^k(\tau_i') = RAT_i^k(t'_{i,0}) \subseteq RAT_i^{k-1}(t'_{i,0}) = RAT_i^{k-1}(t'_{i,1})$ 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 order-0 rationally 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 and Conclusions

The total reward given to the players by our mechanism is upperbounded 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 $G^k - \varepsilon$.

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 leverage the players' higher-order set-theoretic beliefs. Indeed, attractive social choice correspondences defined over such beliefs may be studied and successfully implemented.

Acknowledgements

We have greatly benefitted from the comments of a wonderful anonymous referee.

The first two authors thank Shafi Goldwasser, Andrew Lo, and Ron Rivest for their comments. The third author wishes to thank Joseph Halpern for introducing him to the area of epistemic game theory, and for hours and hours of enlightening discussions about it. This work has been partially supported by ONR Grant No. N00014-09-1-0597.

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.

[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. Hierarchies of beliefs andcommonknowledge. *Journal of Economic Theory*, Vol. 59, pp. 189-198, 1993.

[14] J. Chen and S. Micali. Mechanism Design with Set-Theoretic Beliefs. *Symposium on Foundations of Computer Science (FOCS)*, pp. 87-96, 2011.

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

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

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

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

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

[20] I. Gilboa and D. Schmeidler. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, Vol. 18, pp. 141-153, 1989.

[21] 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.

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

[23] 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.

[24] 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.

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

[26] 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.

[27] J. Weinstein and M. Yildiz. A Structure Theorem for Rationalizability with Application to Robust Predictions of Refinements. *Econometrica*, 75(2), pp. 365-400, 2007.