Computer Science and Artificial Intelligence Laboratory

Technical Report

MIT-CSAIL-TR-2012-017 June 22, 2012

Epistemic Implementation and The Arbitrary-Belief Auction

Jing Chen, Silvio Micali, and Rafael Pass

massachusetts institute of technology, cambridge, ma 02139 usa — www.csail.mit.edu

Epistemic Implementation and The Arbitrary-Belief Auction

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

June 21, 2012

Abstract

In settings of incomplete information we put forward an epistemic framework for designing mechanisms that successfully leverage the players' arbitrary higher-order beliefs, even when such beliefs are totally wrong, and even when the players are rational in a very weak sense. Following Aumann [5], we consider a player i rational if he uses a pure strategy s_i such that no alternative pure strategy s_i' performs better than s_i in every world i considers possible, and consider him *order-k rational* if he is rational and believes that all other players are order-$(k-1)$ rational. We then introduce an iterative deletion procedure of dominated strategies and use it to precisely characterize the strategies consistent with the players being order-k rational.

We exemplify the power of our framework in single-good auctions by introducing and achieving a new class of revenue benchmarks, defined over the players' arbitrary beliefs, that can be much higher than classical ones, and are unattainable by traditional mechanisms. Namely, we exhibit a mechanism that, for every $k \geq 0$ and $\varepsilon > 0$ and whenever the players are *order-$(k+1)$ rational*, guarantees revenue $\geq G^k - \varepsilon$, where G^k is the second highest belief about belief about ... (k times) about the highest valuation of *some* player, even when such a player's identity is not precisely known. Importantly, our mechanism is *possibilistic interim individually rational*. Essentially this means that, based on his beliefs, a player's utility is non-negative not in expectation, but in each world he believes possible.

We finally show that our benchmark G^k is so demanding that it separates the revenue achievable with order-k rational players from that achievable with order-$(k+1)$ rational ones. That is, no possibilistic interim individually rational mechanism can guarantee revenue $\geq G^k - c$, for any constant $c > 0$, when the players are only order-k rational.

1 Introduction

Implementation in settings of *incomplete* information are currently limited by
(1) the inability of leveraging the players' arbitrary higher-order beliefs, and
(2) the strong assumption that all players are expected-utility maximizers.

For our first point, in settings of incomplete information the players are free to form *arbitrary* beliefs, of *arbitrary* order, about the (payoff) types of their opponents. A player's order-0 beliefs are his own type; his order-1 beliefs are his beliefs about his opponents' types; his order-2 beliefs are his beliefs about his opponents' order-1 beliefs; and so on. We refer to beliefs of order 2 or higher as *higher-order beliefs*. Higher-order beliefs can deeply affect a player's strategic decisions, yet such beliefs are not leveraged by classical mechanisms. In particular, dominant-strategy mechanisms only leverage order-0 beliefs, and current Bayesian mechanisms only order-1 beliefs.[1] Also the recent non-Bayesian mechanisms of [12] only leverage order-1 beliefs.

As for our second point, the assumption that all players are expected-utility maximizers is widely relied upon, but is also widely recognized to be very strong. Indeed, a lot of empirical evidence points out that real players may not be so rational (see, in particular, the famous Allais Paradox [2]).

Accordingly, a mechanism designer who ignores the players' higher-order beliefs limits the set of achievable outcomes; and one who assumes that all players are expected-utility maximizers limits the applicability of his mechanisms. We believe these limitations to be serious, but not intrinsic. We note that epistemic game theory has long studied higher-order beliefs as well as weaker notions of rationality. We thus wish to utilize this body of knowledge in order to alleviate the above current limitations of implementation.

Our contribution is three-fold. First, we put forward an epistemic framework enabling mechanism designers to leverage arbitrary higher-order beliefs, *even* when the players are rational in a very weak sense and their beliefs totally wrong.

Second, we apply our framework to single-good auctions. Namely, we construct a mechanism that, whenever the players are order-$(k+1)$ rational, guarantees a very ambitious revenue benchmark, G^k, defined over the players' order-k beliefs.

Finally, we prove that each additional order of rationality provides additional power to implementation. Specifically, we prove that, *for every* k, the benchmark G^k cannot be guaranteed when the players are at most order-k rational. As far as we know, previously it was not even clear whether such a separation existed for —say— order-4 and order-5 rationality.

[1] Dominant-strategy mechanisms leverage only the players' order-0 beliefs *by definition*, because each player's best strategy is to report his own true type, no matter what his opponents might do. Bayesian mechanisms that assume a *common prior* do not even envisage the players having *arbitrary* higher-order beliefs, but require their higher-order beliefs to be inferrable from those of order 1. Many Bayesian mechanisms in the literature actually make an additional assumption: namely, the players' payoff types are independently distributed.

2 Intuitive Explanation of Our Contributions

2.1 An Epistemic Approach to Implementation

Our model of the players' higher-order beliefs follows the *possible worlds* model proposed independently by Aumann [4] and Kripke [18]. Their model has been widely used in the literature of epistemic game theory. But in most of the literature, the players' beliefs about types and beliefs about strategies are treated together. In our approach, we find it important to separate these two kinds of beliefs.

Type Frameworks and Game Frameworks We specify the players' (payoff) types, and their arbitrary-order beliefs about types, via what we call a *type framework*. We succinctly represent such a framework via a *Kripke structure*.[2] Roughly speaking, a type framework specifies a set Ω of possible states of the world. Each state $\omega \in \Omega$ further specifies, for each player i, a payoff type for i and the set of states that i considers possible, that is, i's belief at ω.

We then define the notion of a *game framework* from that of a type framework by "enriching it with beliefs about strategies."[3] Again, we do so via Kripke structures. (Our formalization can be considered as a *possibilistic* variant of the notion of a structure in [15].)

Epistemic Social Choice Correspondences Separately formalizing type frameworks enables us to generalize the traditional notion of a social choice correspondence. In essence, an *epistemic social choice correspondence* is a function mapping not only the players' types, but also their beliefs about types, to outcomes. Such correspondences enable a mechanism designer to enlarge dramatically the set of objectives she may hope to achieve.

Aumann Rationality Following Aumann [5], we take a very conservative approach to defining rationality. Roughly speaking, a player i is *rational* if he uses a pure strategy s_i such that, for every alternative pure strategy s_i', there exists *some* state of the world that i considers possible, where s_i performs as well as s_i'. Thus, we do not need to assume distributions over states: it suffices to work with "possibilistic" beliefs. This notion of rationality is significantly weaker than expected-utility maximization.[4]

Higher orders of rationality are naturally defined as follows: player i is *order-k rational* if he is rational, and believes that all other players are order-$(k-1)$ rational.

Our Notion of Implementation The notion of an implementation is both closely related to that of a solution concept and very demanding. Essentially, it requires that a desired property hold not just at some strategy profiles specified by a given solution concept C, but at *all* of them.

[2] For a good exposition of Kripke structures, see [13].

[3] Traditionally, what we call a game framework is referred to as a "type structure." Since we wish to treat types and strategies separately but in a uniform manner, we have adopted the term "framework."

[4] Expected-utility maximizers are certainly Aumann rational, but whereas there are many experiments showing that people do not act as expected-utility maximizers, we are not aware of any experiments showing that people do not act rationally according to Aumann's notion.

We put forward a class of implementation notions that are conceptually simple and suitable for epistemic social choice correspondences. Essentially,

for an integer $k \geq 0$, our corresponding solution concept consists of all profiles of order-k rational strategies.

Our notions of implementation

- do *not* depend on common belief of rationality (a very strong assumption);
- do *not* require any consistency about the beliefs of different players; and
- are *"closed under Cartesian product."* That is, each underlying solution concept C is of the form $C_1 \times \cdots \times C_n$, where each C_i is a subset of the pure strategies of player i.

This closure property is important from a purely epistemic perspective because it overcomes the "epistemic criticism" of the Nash equilibrium concept, see [7, 6, 3]. 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 strategies) may easily yield undesired outcomes.

2.2 Our Characterization of Order-k Aumann Rationality

We characterize order-k rationality via a new iterative strategy-deletion procedure. Very roughly speaking, for every state ω of the world in the type framework, we keep a set of possible pure strategies for each player. In each iteration, a strategy is removed if it is strictly dominated by some other *pure* strategy in *every* state that i considers possible at ω. Our elimination procedure is actually defined Section 5, where we also formalize (as Theorem 1) and prove the following

> **Characterization Result** (Informal statement.) *For all players i and $k \geq 0$, a strategy of i is order-k rational if and only if it survives k rounds of iterated elimination.*[5]

Note that our characterization of order-k rationality is similar in spirit to the traditional one of *rationalizability* in normal-form games (see [11, 19, 8, 20]). Also note, however, that two main differences exist. First, our characterization applies to games of *incomplete* information. Second, our characterization relies on Aumann's weaker notion of rationality [5], rather than the traditional stronger notion of expected-utility maximization. (This is why we only consider domination by *pure* strategies.)

2.3 An Auction Leveraging Arbitrary Higher-Order Beliefs

We apply our epistemic framework to single-good auctions in a private-value setting[6]. Specifically, we put forward and implement the following revenue benchmarks.

[5] Strategy profiles surviving all iterations are further characterized by common belief of rationality.

[6] Our actual result applies also to the more general setting of *interdependent values*.

Higher-Order-Belief Revenue Benchmarks For every k, we recursively define a revenue benchmark G^k on the players' order-k (possibilistic) beliefs. For simplicity, below we informally define only G^0, G^1 and G^2.

- Let $g_i^0 = \theta_i$ for each player i, where θ_i denotes i's true valuation for the good. (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 v_i such that player i believes that, no matter what the true valuation profile θ may be, there always exists some player j (whose identity need not be known to i) such that $g_j^0 \geq v_i$.
  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 v_i such that player i believes that there always exists some player j (whose identity need not be known to i) such that $g_j^1 \geq v_i$.
  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, the revenue benchmark achieved by the second-price mechanism. As it will become clear from the formal definitions, G^1 coincides with the second-belief benchmark of [12]; $G^0 \leq G^1 \leq G^2 \leq \cdots$; and the gap between G^k and G^{k+1} can be arbitrarily large.

Also note that, since we allow them to be arbitrary, the players' beliefs can be *totally wrong*. In this case, G^k may, for $k > 0$, 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 at least for 200, and player 2 believes that player 1 values it for 300. Then $G^1 = 200$ in this example. However, if all players' beliefs are "correct at every order" (see our technical sections for a proper definition), then every G^k lies in between the highest and second highest true valuation, and can be arbitrarily close to either.

A Single Mechanism Leveraging All Belief Orders We prove the following.

> **Possibility Result** (Informal statement.) *For every $\varepsilon > 0$ there exists an auction mechanism M_ε that, for every k, when run with order-$(k+1)$ rational players, always generates revenue $\geq G^k - \varepsilon$.*

This result is formalized as Theorem 2 in Section 6.5.

Notice that our possibility 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. By running our M_ε we are *automatically* guaranteed to get revenue $\geq G^0 - \varepsilon$ if our 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.

Notice too that, before this result, no interesting social choice correspondence was known to be implementable with —say— order-3 rationality. Prior mechanisms (e.g., the one of [12]) required at most order-2 rationality, or common belief of rationality, but nothing in between.

Roughly speaking, our mechanism is a second-price auction with a reserve price. The mechanism sets the reserve price via information provided by the players themselves, based

on their beliefs. The players are actually paid by the mechanism to provide this information. The idea of buying information from the players is not new. (In particular, it is used by the auction mechanism of [12].) We are not aware, however, of any mechanism where higher-order beliefs are being bought. In some sense, in our mechanism the seller is paying to hear even the *faintest rumors*.

Finally, 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 greatly exceeds the highest valuation, then at least one player has negative utility. This is so because in this case our mechanism generates revenue higher than the highest valuation. Nonetheless, our mechanism is *possibilistic interim individually rational:* that is, as formally defined later, every player *believes* that his utility will be non-negative. Thus every player 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.)

2.4 The Necessity of the Right Rationality Order

We prove that order-$(k+1)$ rationality is necessary to guarantee benchmark G^k. Namely,

> **Impossibility Result** (Informal statement.) *For every $c > 0$ and every k, there is no possibilistic interim individually rational auction mechanism that guarantees revenue $\geq G^k - c$ if the players are only order-k rational.*

This result is formalized as Theorem 4, proved in Section 7.

3 Additional Related Work

As already mentioned, we leverage the players' beliefs in a non-Bayesian setting, and our notion of implementation very different from implementation in dominant strategies.

Weinstein and Yildiz [21] also study iterated elimination and rationalizability based on the players' arbitrary-order beliefs, but in a Bayesian setting.

Although we avoid relying on common belief of rationality, our notions and mechanism can be based on it too, again in a setting of *incomplete* information. Traditionally, implementation under common belief of rationality has been studied for settings of complete information (in particular, see [1] and [14]).

The literature on robust mechanism design, as initiated by Bergemann and Morris [9], is close in spirit to our work. Robust mechanism design too aims at relaxing the common prior assumption, and directly considers richer type spaces to model the players' higher-order beliefs. But the questions it studies are quite different from ours. In particular, robust mechanism design has been used to provide additional justification for implementation in dominant strategies. Also, [9] and subsequent papers still define social choice correspondences over the players' payoff types only (rather than their arbitrary-order beliefs). Yet, Bergemann and Morris [10] explicitly point out that such restricted social choice correspondences cannot represent revenue maximizing allocations. Indeed, our results establish that higher revenue benchmark can be defined and achieved, if considering players' higher-order beliefs when defining the benchmark.

Chen and Micali [12] 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 [12] 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 implies implementation in undominated strategies [17], but not vice versa.

4 The Epistemic Framework

We build our epistemic framework in four steps. First, we formalize *game frameworks* via an intermediate concept, the *type framework*, through which we shall focus on just the players' beliefs about types. Next, we formalize *epistemic contexts* and *epistemic social choice correspondences*, so as to enable a mechanism designer to express his desired outcomes based on both the players' types and their beliefs about types. Then, we formalize Aumann rationality, of any order. Finally, we complete our epistemic framework with a very natural and robust notion of implementation.

We start by recalling some classical "belief-free" notions, so as to establish the following

Basic Notation

- An **environment** is a 4-tuple $(n, \mathcal{O}, \Theta, u)$, where $[n]$ is the set of players; $\mathcal{O}$ the set of outcomes; $\Theta = \Theta_1 \times \cdots \times \Theta_n$ the set of all possible (payoff) type profiles;[7] and u the profile of utility functions, each mapping $\Theta \times \mathcal{O}$ to $\mathbb{R}$, the set of reals.
- The profile of true types is consistently denoted by θ. In an environment $E = (n, \mathcal{O}, \Theta, u)$, $\theta \in \Theta$. We refer to such a pair (E, θ) as a **basic context** for E.
- A **mechanism** M for an environment $E = (n, \mathcal{O}, \Theta, u)$ consists of a set of (pure) strategy profiles, $S = S_1 \times \cdots \times S_n$, and an outcome function (as usual also denoted by M) from S to $\mathcal{O}$ —or to $\Delta(\mathcal{O})$ if M is probabilistic.[8].
- A **basic game** consists of a basic context and a mechanism for the same environment. For every n-player game Γ, we automatically let $\Theta(\Gamma)$ denote the set of type profiles of Γ, $S_i(\Gamma)$ the set of pure strategies for player i in Γ, and $u_i(\Gamma)$ player i's utility function in Γ. Whenever no ambiguity may rise about the game in question, we "drop Γ" and more simply let Θ, S_i and u_i refer, respectively, to $\Theta(\Gamma)$, $S_i(\Gamma)$, and $u_i(\Gamma)$.

[7] When we mention the "type" of some player, we always mean his payoff type —that is, we are distinguishing the players' payoff types from their beliefs.

[8] As usual, $\Delta(A)$ denotes the set of probability distributions over set A.

4.1 Type and Game Frameworks

For simplicity, we consider only *finite* frameworks. As already mentioned, we model the players' beliefs in a set-theoretic way, and thus have no need to assume probability distributions.

Definition 1. *Let $E = (n, \mathcal{O}, \Theta, u)$ be an environment. A **type framework** $\mathcal{V}$ for (n, Θ) consists of a finite set of states Ω; a function $\mathbf{v} : \Omega \to \Theta$; and, a profile of functions $\mathcal{P}$, where $\mathcal{P}_i : \Omega \to 2^\Omega$ for each player i, such that $\forall \omega \in \Omega$,*

1. *$\mathcal{P}_i(\omega) \subseteq \{\omega' \in \Omega : \mathbf{v}(\omega')_i = \mathbf{v}(\omega)_i\}$; and*
2. *$\mathcal{P}_i(\omega) \subseteq \{\omega' \in \Omega : \mathcal{P}_i(\omega') = \mathcal{P}_i(\omega)\}$.*

*The players' beliefs in $\mathcal{V}$ are **correct** (at every order) if, $\forall i$ and $\forall \omega \in \Omega$, we have $\omega \in \mathcal{P}_i(\omega)$.*

In a type framework $\mathcal{V} = (\Omega, \mathbf{v}, \mathcal{P})$, Ω represents the possible states of the world. Our definition does not say anything about the "true state of the world" nor about the "actual beliefs of the players" (these will be considered in the notion of an epistemic context). For each $\omega \in \Omega$, if the true state of the world were ω, then $\mathcal{V}$ defines an infinite hierarchy of beliefs. Namely, for each player i

- the states in $\mathcal{P}_i(\omega)$ are those that i believes possible;
- $\mathbf{v}(\omega)_i$ is the true type of i;
- for each player j, $\{\mathbf{v}(\omega')_j : \omega' \in \mathcal{P}_i(\omega)\}$ represents what types i believes j may have;
- for each pair of players j and k, $\{\mathbf{v}(\omega'')_k : \omega' \in \mathcal{P}_i(\omega), \omega'' \in \mathcal{P}_j(\omega')\}$ represents what types i believes j believes k may have;
- etc.

Conditions 1 and 2 in Definition 1 express that i has the same type and beliefs in every state of the world he believes possible. However, a type framework does not impose any consistency requirements among the beliefs of different players. Indeed, a player may have totally wrong beliefs about another player's beliefs. For instance, in a single-good auction, player i may believe that player j's valuation for the good is greater than 100, whereas player j may believe that player i believes that j's valuation is less than 10.

Note also that $\mathbf{v}$ is a function from Ω to Θ, rather than a profile of functions, where each $\mathbf{v}_i$ maps Ω to Θ_i. This choice enables us to deal with interdependent-type settings as well.

Graphical Representation We find it useful to represent a type framework as a directed graph with labeled nodes and edges. In such a graph:

- a node (drawn as a circle in this paper) represents a state;
- the label of a node (drawn inside the circle) represents the type profile associated with the corresponding state by the function $\mathbf{v}$;
- an edge is labeled by a player, and each node has at least one out-going edge labeled by i for each player i;
- for all states ω and ω', there is an edge with label i from the node of ω to the node of ω' if and only if $\omega' \in \mathcal{P}_i(\omega)$.

To save space, if two edges have the same starting nodes and end nodes, we draw them as a single edge with multiple labels.

As an example, Figure 1 represents a type framework for $n = 2$ and $\Theta = \{0, 1, \ldots, 10\} \times \{0, 1, \ldots, 10\}$ and with 6 states. From this figure we can see that at state ω: the players' types are 3 and 7; and player 1 believes that only two states are possible. In one of them, player 2's type is 4 and he believes that player 1's type is either 0 or 8. In the other, player 2's type is 5 and he believes that player 1's type is 3.

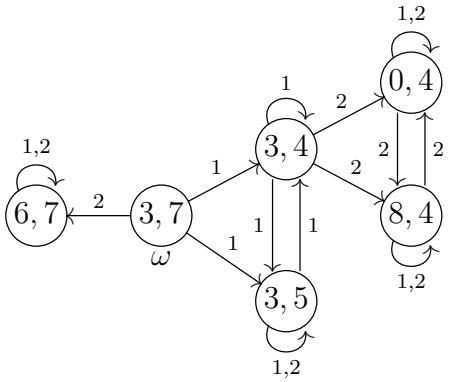


Figure 1: A type framework represented by a directed graph.

Definition 2. *Let M be a mechanism, for an environment $(n, \mathcal{O}, \Theta, u)$, with strategy space S. A **game framework** $\mathcal{M}$ for M consists of a type framework $\mathcal{V} = (\Omega, \mathbf{v}, \mathcal{P})$ for (n, Θ) and a function $\mathbf{s} : \Omega \to S$, such that $\forall i$ and $\forall \omega \in \Omega$, $\mathcal{P}_i(\omega) \subseteq \{\omega' \in \Omega : \mathbf{s}(\omega')_i = \mathbf{s}(\omega)_i\}$.*

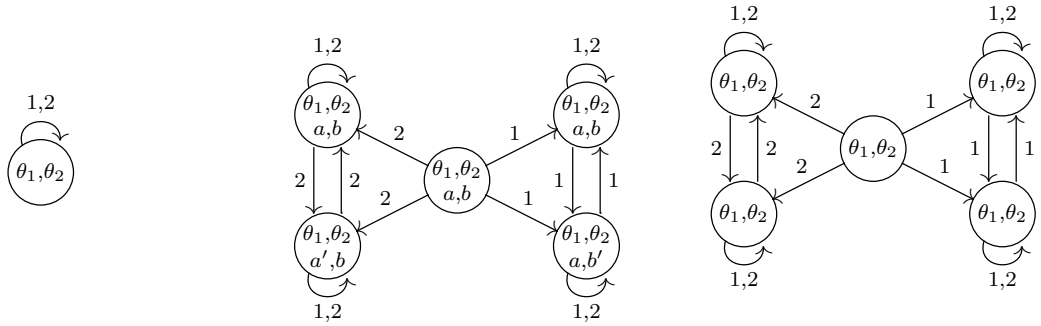
*Such an $\mathcal{M}$ is **consistent with** a type framework $\mathcal{V}' = (\Omega', \mathbf{v}', \mathcal{P}')$ for (n, Θ) if there exists a function $\psi : \Omega \to \Omega'$ such that $\forall \omega \in \Omega$, $\mathbf{v}(\omega) = \mathbf{v}'(\psi(\omega))$ and $\psi(\mathcal{P}_i(\omega)) = \mathcal{P}'_i(\psi(\omega))$ $\forall i$. We refer to such a ψ as a **consistency mapping**.*

To emphasize the underlying type framework $\mathcal{V}$, we may write $\mathcal{M} = (\Omega, \mathbf{v}, \mathcal{P}, \mathbf{s})$ instead of $\mathcal{M} = (\mathcal{V}, \mathbf{s})$. In a game framework $\mathcal{M}$, the constraint of the function $\mathbf{s}$ expresses that, for each state ω, if ω were the true state of the world, then each player i would know his own strategy at ω.

Note that the notion of a game framework can be defined directly, without defining type frameworks first. But we find it important to single out the players' higher-order beliefs about (payoff) types via type frameworks, so that we are able to talk about the pre-existing information in an implementation problem. Indeed, the players may form beliefs about their opponents' true types *before* a designer introduces a mechanism (and thus strategies) into the picture. the notion of consistency captures that, the introduction of a mechanism does not cause the players to change their beliefs about *types*, but causes them to form additional beliefs about *strategies*.

Graphically a game framework $\mathcal{M}$ can be represented by a directed graph as before, except that a node now has an additional label, corresponding to the strategy profile specified by $\mathbf{s}$. Thus, if $\mathcal{M} = (\mathcal{V}, \mathbf{s})$, then the graphical representation of the type framework $\mathcal{V}$ can be obtained by "removing" the strategy label from that of $\mathcal{M}$.

Example Figure 2 provides an elementary game framework for a 2-player mechanism M with strategy space $\{a, a'\} \times \{b, b'\}$, consistent with an elementary type framework $\mathcal{V}'$. The underlying type framework $\mathcal{V}$ in $\mathcal{M}$ is then illustrated in Figure 2c. It is immediate to see that the consistency mapping ψ is the one that maps all states of $\mathcal{M}$ to the only state of $\mathcal{V}'$. Indeed, under such mapping the types are preserved and "the belief function and ψ commute."



(a) Type framework $\mathcal{V}'$ (b) Game framework $\mathcal{M}$ (c) Type framework $\mathcal{V}$

Figure 2: A Trivial Example of Consistency

4.2 Epistemic Contexts and Social Choice Correspondences

When a designer considers which mechanism to choose, we must model what pre-exists the mechanism and is available to the designer, and what pre-exists the mechanism and is not available to him. We take the first piece of information to consist of the environment, and the second of the actual types of the players and the beliefs (of any order) that each player has about the types of his opponents. We now formalize the latter piece of information.[9]

Definition 3. *An **epistemic context** C for an environment $E = (n, \mathcal{O}, \Theta, u)$ consists of two profiles, $\boldsymbol{\mathcal{V}}$ and $\boldsymbol{\omega}$, where, for all i, $\boldsymbol{\mathcal{V}}_i = (\Omega^{(i)}, \mathbf{v}^{(i)}, \mathcal{P}^{(i)})$ is a type framework and $\boldsymbol{\omega}_i \in \Omega^{(i)}$.*

*The only information about C of a player i consists of $\boldsymbol{\mathcal{V}}_i$, $\mathbf{v}^{(i)}(\boldsymbol{\omega}_i)_i$, and $\mathcal{P}_i^{(i)}(\boldsymbol{\omega}_i)$. In C, i's **true type** is $\theta_i \triangleq \mathbf{v}^{(i)}(\boldsymbol{\omega}_i)_i$; and i's **utility** for an outcome $o \in \mathcal{O}$ is $u_i(\mathbf{v}^{(i)}(\boldsymbol{\omega}_i), o)$.*

To emphasize E, we may write $C = (n, \mathcal{O}, \Theta, u, \boldsymbol{\mathcal{V}}, \boldsymbol{\omega})$ instead of $C = (\boldsymbol{\mathcal{V}}, \boldsymbol{\omega})$.

Note that such an epistemic context generalizes the basic context $(n, \mathcal{O}, \Theta, u, \theta)$ by adding arbitrary and independent beliefs for every player.

Note also that in an interdependent-type setting, i may not be able to compute his own utility, because he does not know $\mathbf{v}^{(i)}(\boldsymbol{\omega}_i)$. However, he is able to compute his utility in any state he believes possible. In a private-type setting, however, i's utility for an outcome o can actually be written as $u_i(\mathbf{v}^{(i)}(\boldsymbol{\omega}_i)_i, o)$, and i is always able to compute it.

Definition 4. *A (deterministic) **epistemic social choice correspondence** for an environment $E = (n, \mathcal{O}, \Theta, u)$ is a function from the set of epistemic contexts for E to $2^{\mathcal{O}}$.*

[9]Of course, one may consider that (part of) the epistemic context is also available to the designer, but we aim at designing mechanisms in a way that is as detail-free as possible (the Wilson's doctrine [22]).

Recall that a classical (deterministic) social choice correspondence f maps Θ to $2^{\mathcal{O}}$. Thus any classical social choice correspondence can be simulated by an epistemic one, but not vice versa. Therefore, by allowing a mechanism designer to define his desired outcomes over the players' beliefs rather than their types only, we provide him with a larger set of "targets" to choose from.

Note that to ensure that epistemic social choice correspondences are well defined before the mechanism is chosen, we have "disentangled" type frameworks from game frameworks and define epistemic contexts via the former.

Together with a mechanism M, again common knowledge to the players, an epistemic context C yields an epistemic game $\Gamma = (C, M)$. The players' type frameworks of such a Γ are those of C, and a game framework for Γ is one for M. From now on, when talking about contexts and games, we always mean epistemic contexts and epistemic games.

4.3 Rationality

Notation Let Γ be a game, $\mathcal{M} = (\Omega, \mathbf{v}, \mathcal{P}, \mathbf{s})$ a game framework for Γ, i a player, s_i a strategy of i, ω a state in Ω, and ϕ and ϕ' two statements Then we use the following symbols to ease our discussion:

- $\neg\phi$ stands for the negation of ϕ;
- $\phi \wedge \phi'$ for the conjunction of ϕ and ϕ';
- *true* for the tautological statement;
- RAT_i for the statement "i is rational";
- RAT_i^k for "i is order-k rational", for each $k \geq 0$;
- $play_i(s_i)$ for "i uses strategy s_i";
- $B_i(\phi)$ for "i believes that ϕ holds";
- $(\mathcal{M}, \omega) \models \phi$ for "ϕ holds at $(\mathcal{M}, \omega)$"; and
- $(\mathcal{M}, \omega) \not\models \phi$ for "ϕ does not hold at $(\mathcal{M}, \omega)$".

We define

- $RAT_i^0 \triangleq true$;
- $RAT_i^{k+1} \triangleq RAT_i \wedge B_i(\wedge_{j \neq i} RAT_j^k)$, for each $k \geq 0$;[10]
- $(\mathcal{M}, \omega) \models play_i(s_i)$ if and only if $s_i = \mathbf{s}(\omega)_i$;
- $(\mathcal{M}, \omega) \models \neg\phi$ iff $(\mathcal{M}, \omega) \not\models \phi$;
- $(\mathcal{M}, \omega) \models \phi \wedge \phi'$ iff $(\mathcal{M}, \omega) \models \phi$ and $(\mathcal{M}, \omega) \models \phi'$; and
- $(\mathcal{M}, \omega) \models B_i(\phi)$ iff $\mathcal{P}_i(\omega) \subseteq \{\omega' \in \Omega : (\mathcal{M}, \omega') \models \phi\}$.

Now we turn to defining Aumann rationality and order-k rationality.

Definition 5. *Let Γ be a game, $\mathcal{M} = (\Omega, \mathbf{v}, \mathcal{P}, \mathbf{s})$ a game framework for Γ, ω a state in Ω, i a player, and s_i a strategy of i. Then*

- *s_i is **order-**0 **rational** and i is **order-**0 **rational at** $(\mathcal{M}, \omega)$, $(\mathcal{M}, \omega) \models RAT_i^0$;*

[10] That is, i is order-$(k+1)$ rational if and only if he is rational and believes that all other players are order-k rational (thus $RAT_i^1 = RAT_i$).

- s_i *is* ***rational at*** $(\mathcal{M}, \omega)$ *if for every strategy* s_i' *of* i*, there exists* $\omega' \in \mathcal{P}_i(\omega)$ *such that*

$$u_i(\mathbf{v}(\omega'), (s_i, \mathbf{s}(\omega')_{-i})) \geq u_i(\mathbf{v}(\omega'), (s_i', \mathbf{s}(\omega')_{-i}));$$

- i *is* ***rational at*** $(\mathcal{M}, \omega)$*,* $(\mathcal{M}, \omega) \models RAT_i$*, if* $\mathbf{s}(\omega)_i$ *is rational at* $(\mathcal{M}, \omega)$*;*
- *for every integer* $k \geq 1$*,* s_i *is* ***order-*** k ***rational at*** $(\mathcal{M}, \omega)$ *if* s_i *is rational at* $(\mathcal{M}, \omega)$ *and* $(\mathcal{M}, \omega) \models B_i(\wedge_{j \neq i} RAT_j^{k-1})$*; and*
- *for every integer* $k \geq 1$*,* i *is* ***order-*** k ***rational at*** $(\mathcal{M}, \omega)$*,* $(\mathcal{M}, \omega) \models RAT_i^k$*, if* $\mathbf{s}(\omega)_i$ *is order-*k *rational at* $(\mathcal{M}, \omega)$*.*

Notice that rational and order-1 rational are the same thing.

Although we do not need Aumann's notion of common belief of rationality in our paper, it is easy to check that it is consistent with requiring order-k rationality for all k.

We extend the notion of rationality from game frameworks to type frameworks as follows.

Definition 6. *Let* Γ *be a game with environment* $(n, \mathcal{O}, \Theta, u)$*,* $\mathcal{V} = (\Omega, \mathbf{v}, \mathcal{P})$ *a type framework for* (n, Θ)*,* ω *a state in* Ω*,* i *a player,* s_i *a strategy of* i*, and* k *a non-negative integer. Then* s_i *is* ***rational*** *(respectively,* ***order-*** k ***rational****)* ***at*** $(\mathcal{V}, \omega)$*, if there exists a game framework* $\mathcal{M} = (\Omega', \mathbf{v}', \mathcal{P}', \mathbf{s})$ *for* Γ*, consistent with* $\mathcal{V}$ *under some consistency mapping* ψ*, and a state* $\omega' \in \Omega'$*, such that:*

$$\psi(\omega') = \omega, \quad \mathbf{s}(\omega')_i = s_i, \quad \text{and} \quad i \text{ is rational (respectively, order-}k \text{ rational) at } (\mathcal{M}, \omega').$$

Finally, to be rational for a given game, a strategy must be rational relative to the actual context (type framework and world state) of the corresponding player in the game.

Definition 7. *Let* $\Gamma = ((\boldsymbol{\mathcal{V}}, \boldsymbol{\omega}), M)$ *be a game,* i *a player,* s_i *a strategy of* i*, and* k *a non-negative integer. Then* s_i *is* ***rational*** *(respectively,* ***order-*** k ***rational****) in* Γ *if it is rational (respectively, order-*k *rational) at* $(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)$*.*

A strategy profile s *is* ***rational*** *(respectively,* ***order-*** k ***rational****) if this is so for each* s_i*.*

4.4 Order-k Rational Implementation

Definition 8. *Let* $E = (n, \mathcal{O}, \Theta, u)$ *be an environment,* F *an epistemic social choice correspondence, and* k *a non-negative integer. Then, a mechanism* M ***order-*** k ***rationally implements*** F ***for*** E *if:* $\forall$ *contexts* $C = (\boldsymbol{\mathcal{V}}, \boldsymbol{\omega})$ *for* E*, and* $\forall$ *order-*k *rational strategy profiles* s *in the game* (C, M)*,*

$$M_{E,F}(s) \in F(\boldsymbol{\mathcal{V}}, \boldsymbol{\omega}).$$

Notice that the epistemic context is universally quantified *after* the mechanism. Indeed, in this paper we assume that the environment is the only information available to a mechanism designer. Note too our notion can be easily generalized to "common belief of rationality implementation."

Like for other notions of implementation, ours can be strengthened by requiring that the mechanism satisfies some additional desideratum. The one enjoyed by our auction mechanism of Section 6.2 is a generalization of interim individual rationality.

Relative to the traditional notion of implementation at equilibrium, interim individual rationality requires that, at equilibrium, the utility of every player is non-negative in expectation. Informally speaking, in our setting we instead require that, for each possible epistemic context and each player i, there exists a strategy s_i such that, *according to i's beliefs*, s_i is rational and guarantees i non-negative utility. We refer to such a strategy as "safe."

Definition 9. *A mechanism M for an environment $E = (n, \mathcal{O}, \Theta, u)$ with strategy space S is **possibilistic interim individually rational** if, for every game framework $\mathcal{M} = (\Omega, \mathbf{v}, \mathbf{s}, \mathcal{P})$ for M and every player i, there exists a function $\mathbf{safe}_i : \Omega \to S_i$ such that $\forall \omega \in \Omega$,*

1. *$\mathbf{safe}_i(\omega)$ is rational at $(\mathcal{M}, \omega)$, and*
2. *for every state $\omega' \in \mathcal{P}_i(\omega)$, $u_i(\mathbf{v}(\omega'), (\mathbf{safe}_i(\omega), \mathbf{s}(\omega')_{-i})) \geq 0$.*

Condition 1 implies that, if $(\mathcal{M}, \omega) \models B_i(\wedge_{j \neq i} RAT_j^{k-1})$, i.e., if at state ω player i believes that all other players are order-$(k-1)$ rational, then $\mathbf{safe}_i(\omega)$ is order-k rational at $(\mathcal{M}, \omega)$. Note that, if strategy $\mathbf{safe}_i(\omega)$ exists, then player i will be able to compute it without knowing ω. Indeed, knowing $\mathcal{M}$ and $\mathcal{P}_i(\omega)$ is enough.

5 Characterization for Order-k Rationality

Based on the above definition, checking whether a strategy is order-k rational in a given game may not be easy. Below we characterize order-k rationality via an iterated strategy-deletion procedure that we shall use when analyzing our auction mechanism. Our iterated procedure is related to the classic iterated deletion of strongly dominated strategies, but crucial differences exist. After all, the classic notion applies solely to settings of *complete* information, whereas ours applies to settings of *incomplete* information, which include the former ones as a very special case.

Definition 10. *Let $E = (n, \mathcal{O}, \Theta, u)$ be an environment, M a mechanism for E with strategy space S, i a player, and s_i a strategy in S_i.*

- *Let $\mathcal{V} = (\Omega, \mathbf{v}, \mathcal{P})$ be a type framework for (n, Θ), ω a state in Ω, and S'_{-i} a function from Ω to $2^{S_{-i}}$. Then s_i is **strongly dominated at** $(\mathcal{V}, \omega)$ **with respect to** S'_{-i} if, there exists another strategy s'_i of i such that for all $\omega' \in \mathcal{P}_i(\omega)$ and all $s'_{-i} \in S'_{-i}(\omega')$,*

$$u_i(\mathbf{v}(\omega'), (s'_i, s'_{-i})) > u_i(\mathbf{v}(\omega'), (s_i, s'_{-i})).$$

- *For each non-negative integer k, $NSD_i^k(\cdot, \cdot)$ denotes the **non-strictly-dominated function**, mapping each type framework $\mathcal{V} = (\Omega, \mathbf{v}, \mathcal{P})$ for (n, Θ) and each $\omega \in \Omega$ to a subset of S_i, inductively defined as follows: $NSD_i^0(\mathcal{V}, \omega) = S_i$ and, for $k > 0$, $NSD_i^k(\mathcal{V}, \omega)$ is the set of strategies in $NSD_i^{k-1}(\mathcal{V}, \omega)$ that are not strongly dominated at $(\mathcal{V}, \omega)$ with respect to $NSD_{-i}^{k-1}(\mathcal{V}, \cdot)$, where $NSD_{-i}^{k-1}(\mathcal{V}, \omega) \triangleq \prod_{j \neq i} NSD_j^{k-1}(\mathcal{V}, \omega)$.*
- *s_i **survives k rounds of iterated strong dominance at** $(\mathcal{V}, \omega)$ if $s_i \in NSD_i^k(\mathcal{V}, \omega)$.*
- *s_i **survives iterated strong dominance at** $(\mathcal{V}, \omega)$ if it survives k rounds of iterated strong dominance at $(\mathcal{V}, \omega)$ for all k, that is, if $s_i \in NSD_i^\infty(\mathcal{V}, \omega) \triangleq \bigcap_{k \geq 0} NSD_i^k(\mathcal{V}, \omega)$.*

We use $NSD^k(\mathcal{V},\omega)$ to denote the Cartesian product $\prod_i NSD_i^k(\mathcal{V},\omega)$. When defining $NSD_i^k(\mathcal{V},\omega)$, we remove any strategy s_i that is "strongly dominated by some strategy in S_i". Note that the set $NSD_i^k(\mathcal{V},\omega)$ would be the same if we had required that s_i be dominated by some strategy in $NSD_i^{k-1}(\mathcal{V},\omega)$ instead. Note also that $NSD_i^k(\mathcal{V},\omega)$ is always non-empty, as we only consider finite games.

An immediate consequence of Definition 10 is the following lemma, stated without proof.

Lemma 1. *Strategy s_i is not strongly dominated at $(\mathcal{V},\omega)$ with respect to S'_{-i} if and only if there exists some belief $\mathcal{B}_i$ of i, that is, a subset of $\Theta \times S_{-i}$, such that*

- *$\mathcal{B}_i$ **is consistent with** $\mathcal{P}_i$ **and** S'_{-i} **at** $(\mathcal{V},\omega)$. That is, for any $(v,s'_{-i}) \in \mathcal{B}_i$, there exists $\omega' \in \mathcal{P}_i(\omega)$ such that $v = \mathbf{v}(\omega')$ and $s'_{-i} \in S'_{-i}(\omega')$.*
- *s_i **is rational with respect to** $\mathcal{B}_i$. That is, for every alternative strategy s'_i of i, there exists $(v,s'_{-i}) \in \mathcal{B}_i$ such that $u_i(v,(s_i,s'_{-i})) \geq u_i(v,(s'_i,s'_{-i}))$.*

Lemma 1 is a possibilistic analog of Pearce's lemma [19] which, in probabilistic models, relates best responses and rationality to strong dominance. Note that whereas in the possibilistic case (which is what we consider) this proof is trivial, Pearce's original lemma for the probabilistic case requires additional work.

Another simple but important property of our iterated procedure is the following.

Lemma 2. *For each state ω' in $\mathcal{P}_i(\omega)$ and each $k \geq 0$, $NSD_i^k(\mathcal{V},\omega) = NSD_i^k(\mathcal{V},\omega')$.*

Proof. By the definition of type frameworks, we have $\mathcal{P}_i(\omega') = \mathcal{P}_i(\omega)$. Thus, the definition of strong dominance implies that, for any strategy s_i, s_i is strongly dominated at $(\mathcal{V},\omega)$ with respect to some function S'_{-i} if and only if it is strongly dominated at $(\mathcal{V},\omega')$ with respect to the same S'_{-i}. Further because $NSD_i^0(\mathcal{V},\omega) = NSD_i^0(\mathcal{V},\omega') = S_i$, an easy induction implies that $NSD_i^k(\mathcal{V},\omega) = NSD_i^k(\mathcal{V},\omega')$ for each k. ■

Lemma 2 implies that player i is able to compute $NSD_i^k(\mathcal{V},\omega)$ knowing $\mathcal{V}$ and $\mathcal{P}_i(\omega)$, without knowing ω.

We are now ready to state our characterization of order-k rationality.

Theorem 1. *Let Γ be a game with environment $(n,\mathcal{O},\Theta,u)$, $\mathcal{V} = (\Omega,\mathbf{v},\mathcal{P})$ a type framework for (n,Θ), ω a state in Ω, i a player, s_i a strategy of i, and $k \geq 0$. Then,*

$$s_i \in NSD_i^k(\mathcal{V},\omega) \text{ if and only if } s_i \text{ is order-}k \text{ rational at } (\mathcal{V},\omega).$$

Proof of the "if" direction.

Assuming that s_i is order-k rational at $(\mathcal{V},\omega)$, we prove $s_i \in NSD_i^k(\mathcal{V},\omega)$ by induction on k.

For $k = 0$, the property trivially holds since $NSD_i^0(\mathcal{V},\omega) = S_i(\Gamma)$ by definition.

For $k > 0$, by Definition 6 there exists a game framework $\mathcal{M} = (\Omega',\mathbf{v}',\mathcal{P}',\mathbf{s})$ consistent with $\mathcal{V}$ under some consistency mapping ψ, and a state $\omega' \in \Omega'$, such that $\psi(\omega') = \omega$ and $(\mathcal{M},\omega') \models play_i(s_i) \wedge RAT_i^k$. Expanding out the definition of RAT_i^k, we get

$$(\mathcal{M},\omega') \models play_i(s_i) \wedge RAT_i \wedge B_i(\wedge_{j\neq i} RAT_j^{k-1}).$$

Thus, we have

$$(\mathcal{M}, \omega') \models play_i(s_i) \wedge RAT_i, \tag{1}$$

and that for every $\omega'' \in \mathcal{P}'_i(\omega')$,

$$(\mathcal{M}, \omega'') \models \wedge_{j \neq i} RAT_j^{k-1},$$

which further implies that for each $j \neq i$, $\mathbf{s}(\omega'')_j$ is order-$(k-1)$ rational at $(\mathcal{V}, \psi(\omega''))$. By the induction hypothesis it follows that

$$\forall \omega'' \in \mathcal{P}'_i(\omega') \quad \mathbf{s}(\omega'')_{-i} \in NSD_{-i}^{k-1}(\mathcal{V}, \psi(\omega'')). \tag{2}$$

Accordingly, letting $\mathcal{B}_i \triangleq \{(\mathbf{v}(\psi(\omega'')), \mathbf{s}(\omega'')_{-i}) : \omega'' \in \mathcal{P}'_i(\omega')\}$, we have that:

(a) $\mathcal{B}_i$ is consistent with $\mathcal{P}_i$ and $NSD_{-i}^{k-1}(\mathcal{V}, \cdot)$ at $(\mathcal{V}, \omega)$. This follows from the consistency of $\mathcal{M}$ and $\mathcal{V}$, and Equation 2.

(b) s_i is rational with respect to $\mathcal{B}_i$. This follows from Equation 1.

By Lemma 1 we thus have that s_i is not strongly dominated at $(\mathcal{V}, \omega)$ with respect to $NSD_{-i}^{k-1}(\mathcal{V}, \cdot)$.

Since for every $k' > 0$ and $\omega' \in \Omega$ we have $NSD_{-i}^{k'}(\mathcal{V}, \omega') \subseteq NSD_{-i}^{k'-1}(\mathcal{V}, \omega')$, s_i is not strongly dominated at $(\mathcal{V}, \omega)$ with respect to $NSD_{-i}^{k'}(\mathcal{V}, \cdot)$ for any $k' < k$.

It follows that $s_i \in NSD_i^k(\mathcal{V}, \omega)$, concluding the proof of the "if" direction.

Proof of the "only-if" direction.

By definition, proving this direction is equivalent to proving that, if $s_i \in NSD_i^k(\mathcal{V}, \omega)$, then there exists a game framework $\mathcal{M} = (\Omega', \mathbf{v}', \mathcal{P}', \mathbf{s})$ for Γ, consistent with $\mathcal{V}$ under some consistency mapping ψ, and a state $\omega' \in \Omega'$ such that

$$\psi(\omega') = \omega \quad \text{and} \quad (\mathcal{M}, \omega') \models play_i(s_i) \wedge RAT_i^k.$$

Notice that such $\mathcal{M}$ and ψ may depend on k, ω, i, and s_i.

In fact, we shall prove an even stronger statement. Namely, for each k, there exists a *universal* game framework $\mathcal{M} = (\Omega', \mathbf{v}', \mathcal{P}', \mathbf{s})$ for Γ, consistent with $\mathcal{V}$ under some consistency mapping ψ, such that for *every* $\omega \in \Omega$, $k' \leq k$, player i and strategy s_i,

if $s_i \in NSD_i^{k'}(\mathcal{V}, \omega)$ then there exists a state $\omega^{k'} \in \Omega'$ such that

$$\psi(\omega^{k'}) = \omega \quad \text{and} \quad (\mathcal{M}, \omega^{k'}) \models play_i(s_i) \wedge RAT_i^{k'}, \tag{3}$$

which implies that s_i is order-k' rational at $(\mathcal{V}, \omega)$.

Define $\mathcal{M}$ as follows.

- $\Omega' = \{(s', \omega', k', j) : s' \in NSD^{k'}(\mathcal{V}, \omega'), \omega' \in \Omega, k' \in \{0, \ldots, k\}, j \in [n]\}$;
- $\mathbf{v}'(s', \omega', k', j) = \mathbf{v}(\omega')$;
- $\mathbf{s}(s', \omega', k', j) = s'$; and
- $\mathcal{P}'_i(s', \omega', k', j)$ is defined as follows:

  for $i \neq j$:

  $$\mathcal{P}'_i(s', \omega', k', j) = \{((s'_i, s''_{-i}), \omega'', \max(k'-1, 0), i) : \ s''_{-i} \in NSD_{-i}^{\max(k'-1,0)}(\mathcal{V}, \omega''), \ \omega'' \in \mathcal{P}_i(\omega')\}.$$

for $i = j$:

$$\mathcal{P}'_i(s', \omega', k', j) = \{((s'_i, s''_{-i}), \omega'', k', i) : \ s''_{-i} \in NSD^{k'}_{-i}(\mathcal{V}, \omega''), \ \omega'' \in \mathcal{P}_i(\omega')\}.$$

It is easy to check that $\mathcal{M}$ is a well-defined game framework for Γ. Indeed, the reason we define $\mathcal{P}'_i(s', \omega', k', j)$ differently for $i \neq j$ and $i = j$ is to ensure that a player knows his own beliefs. Also by definition it follows that $\mathcal{M}$ is consistent with $\mathcal{V}$ under the consistency mapping ψ where $\psi(s', \omega', k', j) = \omega'$.

Let us prove by induction over k' that for every $\omega \in \Omega, k' \leq k$, strategy profile $s \in NSD^{k'}(\mathcal{V}, \omega)$, and every player j, we have

$$(\mathcal{M}, (s, \omega, k', j)) \models \wedge_{\hat{j} \neq j} RAT^{k'}_{\hat{j}}. \tag{4}$$

Notice that Equation 3 follows directly from Equation 4: just let $\omega^{k'} = ((s_i, s_{-i}), \omega, k', j)$ with some $j \neq i$ and $s_{-i} \in NSD^{k'}_{-i}(\mathcal{V}, \omega)$.

To prove Equation 4, notice that the base case ($k' = 0$) trivially holds, since $RAT^0_{\hat{j}}$ is defined to be *true* for each player $\hat{j}$. For the induction step ($k' > 0$), arbitrarily fixing $s \in NSD^{k'}(\mathcal{V}, \omega)$, it suffices to show that for every $\hat{j} \neq j$,

$$(\mathcal{M}, (s, \omega, k', j)) \models RAT^{k'}_{\hat{j}};$$

that is,

$$(\mathcal{M}, (s, \omega, k', j)) \models RAT_{\hat{j}} \wedge B_{\hat{j}}(\wedge_{j' \neq \hat{j}} RAT^{k'-1}_{j'}). \tag{5}$$

Because $s_{\hat{j}} \in NSD^{k'}_{\hat{j}}(\mathcal{V}, \omega) \subseteq NSD^{k'-1}_{\hat{j}}(\mathcal{V}, \omega)$, and because $NSD^{k'-1}_{\hat{j}}(\mathcal{V}, \omega) = NSD^{k'-1}_{\hat{j}}(\mathcal{V}, \omega')$ for any $\omega' \in \mathcal{P}_{\hat{j}}(\omega)$ by Lemma 2, we have $s_{\hat{j}} \in NSD^{k'-1}_{\hat{j}}(\mathcal{V}, \omega')$ for any $\omega' \in \mathcal{P}_{\hat{j}}(\omega)$. By the induction hypothesis, for any $\omega' \in \mathcal{P}_{\hat{j}}(\omega)$ and $s'_{-\hat{j}} \in NSD^{k'-1}_{-\hat{j}}(\mathcal{V}, \omega')$, we have

$$(\mathcal{M}, ((s_{\hat{j}}, s'_{-\hat{j}}), \omega', k' - 1, \hat{j})) \models \wedge_{j' \neq \hat{j}} RAT^{k'-1}_{j'}.$$

By the definition of $\mathcal{P}'_{\hat{j}}$, this means that for any state $\hat{\omega} \in \mathcal{P}'_{\hat{j}}(s, \omega, k', j)$, we have

$$(\mathcal{M}, \hat{\omega}) \models \wedge_{j' \neq \hat{j}} RAT^{k'-1}_{j'}.$$

By the definition of $B_{\hat{j}}(\phi)$, it thus follows that

$$(\mathcal{M}, (s, \omega, k', j)) \models B_{\hat{j}}(\wedge_{j' \neq \hat{j}} RAT^{k'-1}_{j'}).$$

Therefore to prove Equation 5 it remains to show that

$$(\mathcal{M}, (s, \omega, k', j)) \models RAT_{\hat{j}}.$$

Since $s \in NSD^{k'}(\mathcal{V}, \omega)$, we have that $s_{\hat{j}}$ is not strongly dominated at $(\mathcal{V}, \omega)$ with respect to $NSD^{k'-1}_{-\hat{j}}(\mathcal{V}, \cdot)$. That is, for every alternative strategy $s'_{\hat{j}}$ of $\hat{j}$, there exists $\omega' \in \mathcal{P}_{\hat{j}}(\omega)$ and $s'_{-\hat{j}} \in NSD^{k'-1}_{-\hat{j}}(\mathcal{V}, \omega')$ such that

$$u_{\hat{j}}(\mathbf{v}(\omega'), (s_{\hat{j}}, s'_{-\hat{j}})) \geq u_{\hat{j}}(\mathbf{v}(\omega'), (s'_{\hat{j}}, s'_{-\hat{j}})). \tag{6}$$

Letting $\omega^{k'} = ((s_{\hat{j}}, s'_{-\hat{j}}), \omega', k'-1, \hat{j})$, by definition we have

$$\omega^{k'} \in \mathcal{P}'_{\hat{j}}(s, \omega, k', j), \quad \mathbf{v}'(\omega^{k'}) = \mathbf{v}(\omega'), \quad \mathbf{s}(s, \omega, k', j)_{\hat{j}} = s_{\hat{j}}, \quad \text{and} \quad \mathbf{s}(\omega^{k'})_{-\hat{j}} = s'_{-\hat{j}}.$$

Combining Equation 6 with the equalities above we have that, for every $s'_{\hat{j}}$ there exists $\omega^{k'} \in \mathcal{P}'_{\hat{j}}(s, \omega, k', j)$ such that

$$u_{\hat{j}}(\mathbf{v}'(\omega^{k'}), (\mathbf{s}(s, \omega, k', j)_{\hat{j}}, \mathbf{s}(\omega^{k'})_{-\hat{j}})) \geq u_{\hat{j}}(\mathbf{v}'(\omega^{k'}), (s'_{\hat{j}}, \mathbf{s}(\omega^{k'})_{-\hat{j}})).$$

Thus $(\mathcal{M}, (s, \omega, k', j)) \models RAT_{\hat{j}}$ by definition, and Equation 5 holds. This concludes the proof of the induction step for Equation 4, and the proof of Equation 3. Therefore the "only-if" direction holds, concluding the proof of Theorem 1. ■

Theorem 1 has the following immediate corollary.

Corollary 1. *Let Γ be a finite game with context $(\boldsymbol{\mathcal{V}}, \boldsymbol{\omega})$, i a player, and $k \geq 0$. Then the set of order-k rational strategies of i is $NSD_i^k(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)$, which is always non-empty.*

Finally, note that we can also characterize common belief of rationality: s_i survives iterated strong dominance at $(\mathcal{V}, \omega)$ if and only if it is common-belief rational at $(\mathcal{V}, \omega)$.

6 The Arbitrary-Belief Mechanism

Let us quickly recall single-good auctions so as to establish the following

Auction Notation A (finite) single-good auction environment $E = (n, \mathcal{O}, \Theta, u)$ where

- $\mathcal{O}$ consists of all pairs (a, P), where $a \in \{0, \ldots, n\}$ and $P \in \mathbb{R}^n$.
  We refer to a as the *allocation*, and to P as the *price profile*. If $a = 0$, then the good is unallocated; if $a > 0$, then the good is allocated to player a. If $P_i \geq 0$, then player i pays P_i to the seller; if $P_i < 0$, the i is paid $-P_i$ by the seller.
- $\Theta \triangleq \{0, 1, \ldots, V\}^n$, for some positive integer V.
  A type in Θ_i is referred to as a possible *valuation*, and V as the *valuation bound*. Accordingly, the true type of player i, θ_i, is referred to i's *true valuation*.
- For all $i \in [n]$, $v \in \{0, \ldots, V\}^n$, and $(a, P) \in \mathcal{O}$,
  $u_i(v, (a, P)) = v_i - P_i$ if $a = i$, and $= -P_i$ otherwise.

Notice that an auction environment is fully specified by just n and V, and thus de facto consist of a pair: $E = (n, V)$.

The *revenue* of an outcome (a, P), denoted by $rev(a, P)$, is $\sum_i P_i$.

As usual, an epistemic context for an auction environment $E = (n, V)$ further specifies a type framework profile $\boldsymbol{\mathcal{V}}$ where each $\boldsymbol{\mathcal{V}}_i = (\Omega^{(i)}, \mathbf{v}^{(i)}, \mathcal{P}^{(i)})$, and a state profile $\boldsymbol{\omega}$ where each $\boldsymbol{\omega}_i \in \Omega^{(i)}$. We denote by $\mathscr{C}_n^V$ the set of all (epistemic) contexts for (n, V). From now on we only deal with single-good auctions, and thus refer to a type framework a "valuation framework".

6.1 The Epistemic Revenue Benchmarks G^k

Recall that an epistemic social choice correspondence for an environment E maps each epistemic context for E to a set of outcomes. Below we instead define an *epistemic revenue benchmark* for an auction environment E, that is, a function b mapping each epistemic context $(\boldsymbol{\mathcal{V}}, \boldsymbol{\omega})$ for E to a real number $b(\boldsymbol{\mathcal{V}}, \boldsymbol{\omega})$. This function is equivalent to the epistemic social choice correspondence mapping each $(\boldsymbol{\mathcal{V}}, \boldsymbol{\omega})$ to the set of auction outcomes whose revenue is at least $b(\boldsymbol{\mathcal{V}}, \boldsymbol{\omega})$. The notion of order-$k$ rational implementation is thus well defined for b.

Definition 11. *Let $\mathcal{V} = (\Omega, \mathbf{v}, \mathcal{P})$ be a valuation framework for an auction environment (n, V). Then, for each player i and each integer $k \geq 0$, we inductively define the function g_i^k as follows: $\forall$ state $\omega \in \Omega$,*

$$g_i^0(\mathcal{V}, \omega) = \min_{\omega' \in \mathcal{P}_i(\omega)} \mathbf{v}(\omega')_i, \text{ and } g_i^k(\mathcal{V}, \omega) = \min_{\omega' \in \mathcal{P}_i(\omega)} \max_{j \in [n]} g_j^{k-1}(\mathcal{V}, \omega') \ \forall k \geq 1.$$

*We refer to $g_i^k(\mathcal{V}, \omega)$ as the **order-k guaranteed value of i at** $(\mathcal{V}, \omega)$.*

We so name $g_i^k(\mathcal{V}, \omega)$ because, if $g_i^k(\mathcal{V}, \omega) \geq c$, then, at state ω, player 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 valuation is at least c.

Note that if the players' beliefs in $\mathcal{V}$ are correct at every order, then for each player i, each state ω, and each $k \geq 0$, $g_i^k(\mathcal{V}, \omega) \leq \max_j \mathbf{v}(\omega)_j$. Note also that player i is able to compute the value of $g_i^k(\mathcal{V}, \omega)$ knowing $\mathcal{V}$ and $\mathcal{P}_i(\omega)$, without knowing ω itself. Indeed, the following claim can be trivially proved by induction.

Claim 1. *Let $\mathcal{V} = (\Omega, \mathbf{v}, \mathcal{P})$ be a valuation framework for an auction environment (n, V). Then $\forall \omega \in \Omega$, $\forall i \in [n]$, and $\forall k \geq 0$, $\mathcal{P}_i(\omega) \subseteq \{\omega' : g_i^k(\mathcal{V}, \omega') = g_i^k(\mathcal{V}, \omega)\}$.*

Next, note that the g_i^k's are non-decreasing in k.

Claim 2. *Let $\mathcal{V} = (\Omega, \mathbf{v}, \mathcal{P})$ be a valuation framework for an auction environment (n, V). Then $\forall \omega \in \Omega$, $\forall i \in [n]$, and $\forall k \geq 1$, $g_i^k(\mathcal{V}, \omega) \geq g_i^{k-1}(\mathcal{V}, \omega)$.*

Proof. By Definition 11 and Claim 1, we have

$$g_i^k(\mathcal{V}, \omega) = \min_{\omega' \in \mathcal{P}_i(\omega)} \max_{j \in [n]} g_j^{k-1}(\mathcal{V}, \omega') \geq \min_{\omega' \in \mathcal{P}_i(\omega)} g_i^{k-1}(\mathcal{V}, \omega') = \min_{\omega' \in \mathcal{P}_i(\omega)} g_i^{k-1}(\mathcal{V}, \omega) = g_i^{k-1}(\mathcal{V}, \omega).$$

Thus Claim 2 holds. □

Finally, because $\mathbf{v}(\omega')_i = \mathbf{v}(\omega)_i$ for each $\omega' \in \mathcal{P}_i(\omega)$, we additionally have

$$g_i^0(\mathcal{V}, \omega) = \mathbf{v}(\omega)_i.$$

We are now ready to define our revenue benchmarks for single-good auctions.

Definition 12. *Let (n, V) be an auction environment and k a non-negative integer. Then the **order-k revenue benchmark**, denoted by G^k, is the function mapping each context $(\boldsymbol{\mathcal{V}}, \boldsymbol{\omega})$ for (n, V) to the second highest value in $\{g_i^k(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)\}_{i \in [n]}$.*

Note that G^k (as for other non-Bayesian revenue benchmarks) is actually well defined for all single-good auction environments.[11] Note also that $G^k(\boldsymbol{\mathcal{V}}, \boldsymbol{\omega}) \geq G^{k-1}(\boldsymbol{\mathcal{V}}, \boldsymbol{\omega})$ for each $k > 0$.

Further note that, because each player knows his own valuation, $G^0(\boldsymbol{\mathcal{V}}, \boldsymbol{\omega})$ always coincides with "the second highest valuation" (which is what the standard second-price auction guarantees). Finally note that for any constant $\varepsilon > 0$, the revenue benchmarks $G^k - \varepsilon$ can be defined from G^k in a straightforward way.

6.2 The Arbitrary-Belief Mechanism $M_{n,V,K,\varepsilon}$

In this section, we construct a finite auction mechanism, $M_{n,V,K,\varepsilon}$, that for every auction environment (n, V), positive integer K, $\varepsilon > 0$, and every $k \in \{0, \ldots, K\}$, order-$(k+1)$ rationally implements $G^k - \varepsilon$. The value K is called the *order bound*, specifying the maximum order of beliefs leveraged by our mechanism. That is, if $K = 99$ then, our mechanism leverages the players' 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. This mechanism is actually uniformly constructed on inputs n, V, K, and ε. Since it does not depend on the rationality order of the players, we refer to it as the *Arbitrary-Belief Mechanism*.

(**Note:** the reliance of our mechanism on the order bound K is not crucial —in fact, if we are willing to make the strategy space of our mechanism infinite, then we do not need K and our mechanism can leverage the players' beliefs up to any order. However, since K can be arbitrarily large, we prefer the current, finite formalization of our mechanism.)

In the Arbitrary-Belief Mechanism, a strategy of a player i has three components: (1) his own identity, for convenience only; (2) a *belief-order* $\ell_i \in \{0, \ldots, K\}$; and (3) a *value* $v_i \in \{0, \ldots, V\}$. The mechanism is of normal form. The players simultaneously announce their chosen strategies, and the mechanism decides the winner of the good and an initial price for every player. (As in the second-price mechanism, this price is 0 when a player does not win the good). The final price of each player consists of his initial price minus a reward determined by evaluating a *reward function* ρ mapping strategy profiles to real numbers.

The mechanism is presented below. Note that 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 $(i, \ell_i, v_i) \in \{i\} \times \{0, \ldots, K\} \times \{0, \ldots, V\}$.*

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.*

[11] Actually, G^k is well defined even when the players do not know their own valuations. In such cases, formally speaking a type framework $\mathcal{V} = (\Omega, \mathbf{v}, \mathcal{P})$ is defined dropping the requirement $\mathcal{P}_i(\omega) \subseteq \{\omega' : \mathbf{v}(\omega')_i = \mathbf{v}(\omega)_i\}$. The only difference is that we no longer have $g_i^0(\mathcal{V}, \omega) = \mathbf{v}(\omega)_i$. Both our possibility result and impossibility result hold for such cases.

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

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

We refer to δ_i as player i's reward, ρ as the reward function, and (a, P) as the final outcome.

Note that our mechanism never leaves the good unsold. Let us now formalize the statement of the possibility result promised in the Introduction.

Theorem 2. *For all auction environments (n, V), positive integers K, and $\varepsilon > 0$, the mechanism $M_{n,V,K,\varepsilon}$ is possibilistic interim individually rational and, for each $k \in \{0, \ldots, K\}$, order-$(k+1)$ rationally implements the revenue benchmark $G^k - \varepsilon$.*

6.3 Variants of Our Mechanism and the Arbitrary-Belief Lemma

Even without formal analysis, it is easy to see that in mechanism $M_{n,V,K,\varepsilon}$ the utility of a player who does not get the good is upper-bounded by the absolute value ε/n. Accordingly, a player who believes that he will not win the good also believes that ε/n is the most he can gain from participating to the mechanism. According to traditional economic theory, this is sufficient motivation: indeed, a utility maximizer i strictly prefers outcome o_1 to outcome o_2 whenever $u_i(o_1) - u_i(o_2) > 0$, even if this difference is very small. However, several researchers (with computer scientists in the lead) worry that an "ε motivation" may not be sufficient. The same concern should of course arise in the second-price mechanism. Indeed, if the player with the second highest valuation believes that he is not going to get the good, then it is rational for him to drop out or bid 0 in the second-price mechanism, in which case the revenue fetched may be much lower than the second-highest valuation.[12]

Avoiding these objections requires increasing the rewards given to the players, but then, of course, the revenue collected will diminish correspondingly! A way to increase motivation that merits attention is rewarding the players not with an *absolute* (amount of money) ε, but with an ε *fraction* (e.g., 1%) of the price at which the good is sold. Properly implementing this approach yields the following revenue guarantee.

Theorem 3. *For all auction environments (n, V), positive integers K, and $\varepsilon \in (0, 1)$, there exists a mechanism $M'_{n,V,K,\varepsilon}$ that is possibilistic interim individually rational and, for each $k \in \{0, \ldots, K\}$, order-$(k+1)$ rationally implements the revenue benchmark $(1-\varepsilon)G^k - \varepsilon$.*

This and other revenue guarantees can be obtained by just changing the reward function ρ of the Arbitrary-Belief Mechanism with one satisfying the following property.

Definition 13. *A reward function $\bar{\rho}$ for $M_{n,V,K,\varepsilon}$ is **proper** if it maps any triple $(2^{nd}v, v_i, \ell_i)$ to a real number so as to satisfy the following conditions:*

(1) for all v_i and ℓ_i, $\bar{\rho}(\cdot, v_i, \ell_i)$ is non-decreasing with $2^{nd}v$;

[12] Preventing this requires offering some form of "ε reward" to the players in the second-price mechanism too, thus reducing its revenue guarantee by ε. Once the playing field is so leveled, it can be seen that our mechanism offers at least the same revenue than the second-price one (since players are always assumed to be order-1 rational and G^0 coincides with the second highest valuation), and sometimes much more (if they are order-k rational and have "suitably high" beliefs).

(2) for all $2^{nd}v$, $\bar{\rho}(2^{nd}v, \cdot, \cdot)$ is strictly increasing with v_i —that is, if $v_i > v_i'$, then for all ℓ_i and ℓ_i', we have $\bar{\rho}(2^{nd}v, v_i, \ell_i) > \bar{\rho}(2^{nd}v, v_i', \ell_i')$;

(3) for all $2^{nd}v$ and v_i, $\bar{\rho}(2^{nd}v, v_i, \cdot)$ is strictly decreasing with ℓ_i.

We denote by $M_{n,V,K,\bar{\rho}}$ the mechanism obtained by replacing ρ with $\bar{\rho}$ in $M_{n,V,K,\varepsilon}$.

Note that the reward function ρ of $M_{n,V,K,\varepsilon}$ is indeed proper (although degenerated, since it does not depend on $2^{nd}v$). Note also that the parameter ε enters the Arbitrary-Belief Mechanism only via the reward function ρ, and thus is no longer relevant in $M_{n,V,K,\bar{\rho}}$.

Let us now state a general lemma easily implying both Theorems 2 and 3.

Lemma 3. ***(Arbitrary-Belief Lemma)*** *For all auction environments (n, V), positive integers K, and proper reward functions $\bar{\rho}$, the mechanism $M_{n,V,K,\bar{\rho}}$ is possibilistic interim individually rational and, $\forall$ context $(n, V, \boldsymbol{\mathcal{V}}, \boldsymbol{\omega}) \in \mathscr{C}_n^V$, $\forall\, k \in \{0, \ldots, K\}$, and $\forall$ order-$(k+1)$ rational strategy profile s, we have*

$$rev(M(s)) = 2^{nd}v - \sum_i \delta_i \quad \text{and} \quad 2^{nd}v \geq G^k(\boldsymbol{\mathcal{V}}, \boldsymbol{\omega}).$$

6.4 Proof of the Arbitrary-Belief Lemma

To prove Lemma 3, we write $M_{n,V,K,\bar{\rho}}$ as M for short. The equality for the revenue of M follows directly from the description of M. To show that M is possibilistic interim individually rational is easy and done at the end. The more complex part is showing that $2^{nd}v \geq G^k(\boldsymbol{\mathcal{V}}, \boldsymbol{\omega})$ for any order-$(k+1)$ rational strategy profile s, which is done below. To ease the discussion, we first provide the following

6.4.1 Over-Simplified Analysis of M

To start with, notice that as long as each player i's announced value v_i is at least $g_i^k(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)$, $2^{nd}v$ is at least $G^k(\boldsymbol{\mathcal{V}}, \boldsymbol{\omega})$ by the definition of G^k. Therefore the main point of the analysis is to show that, for any strategy $s_i = (i, \ell_i, v_i) \in NSD_i^{k+1}(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)$, $v_i \geq g_i^k(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)$.

Assuming $v_i < g_i^k(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)$, it suffices to prove that s_i is strongly dominated by another strategy $\hat{s}_i = (i, \hat{v}_i, \hat{\ell}_i)$ at $(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)$ with respect to $NSD_{-i}^k(\boldsymbol{\mathcal{V}}_i, \cdot)$. To do so, we take $\hat{s}_i$ to be the alleged strategy, that is, $\hat{v}_i = g_i^k(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)$ and $\hat{\ell}_i = \min\{\ell : g_i^\ell(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i) = g_i^k(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)\}$. Because $\bar{\rho}$ is proper, no matter what the other players do, using $\hat{s}_i$ gives player i more reward than using s_i.

But this is not enough to prove the desired domination. Because when $g_i^k(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i) > g_i^0(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)$, in principle there exists the case where, by using s_i player i does not get the good, and by using $\hat{s}_i$ he gets the good and pays a price higher than $g_i^0(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)$. In such a case i's utility could be negative using $\hat{s}_i$, while positive using s_i. The key point here is to show that such a case never occurs according to player i's belief —that is, assuming order-$(k+1)$ rationality, if $g_i^k(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i) > g_i^0(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)$, then player i believes that by using $\hat{s}_i$ he never gets the good (and thus the bigger reward is his pure gain).

To prove this, we do induction on k. The first induction hypothesis is about the value of v_i. By the definition of $g_i^k(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)$, player i believes that there is always some player who believes that ... (k times) that there is always some player who values the good for at

least $g_i^k(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)$. Accordingly, in any state of the world ω' considered possible by player i at $\boldsymbol{\omega}_i$, there exists some player j whose order-$(k-1)$ guaranteed value $g_j^{k-1}(\boldsymbol{\mathcal{V}}_i, \omega')$ is at least $g_i^k(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)$. Because player i is order-$(k+1)$ rational, he believes that all the other players are order-k rational. Accordingly, by induction hypothesis, in player i's belief, at state ω' player j's announced value is at least $g_j^{k-1}(\boldsymbol{\mathcal{V}}_i, \omega')$. If $g_j^{k-1}(\boldsymbol{\mathcal{V}}, \omega') > g_i^k(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)$, of course player i cannot get the good (actually we also need to show that $j \neq i$, but this follows easily from the definition of the g^k's).

What if $g_j^{k-1}(\boldsymbol{\mathcal{V}}_i, \omega') = g_i^k(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)$? To deal with this possibility, we need another induction hypothesis, about the value of ℓ_i. Again because the reward function $\bar{\rho}$ is proper, fixing $\hat{v}_i$, announcing $\hat{\ell}_i$ gives player i a bigger reward than announcing anything greater than $\hat{\ell}_i$. By induction hypothesis, player j above, being order-k rational in i's belief, is going to announce the lowest belief order ℓ' such that $g_j^{\ell'}(\boldsymbol{\mathcal{V}}_i, \omega') = g_j^{k-1}(\boldsymbol{\mathcal{V}}_i, \omega')$. By the definition of the g^k's, it can be proved that ℓ' is at most $\hat{\ell}_i - 1$, that is, $\ell' < \hat{\ell}_i$. From the way how the players' announced triples are ordered, player j is ordered before i, and thus i cannot get the good.

To summarize, if player i believes that some player believes that ... that some player values the good for at least $g_i^k(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)$, then it is always "safe" for i to use his alleged strategy, which gives him the biggest reward without any risk of being over charged. This concludes our simplified analysis. Let us now turn to a

6.4.2 Detailed Analysis of M

Definition 14. *For any two pairs of non-negative integers (ℓ, v) and (ℓ', v'), we say that (ℓ, v) is **greater than** (ℓ', v'), written as $(\ell, v) \succ (\ell', v')$, if: either $v > v'$, or $v = v'$ and $\ell < \ell'$. We say that (ℓ, v) is **great than or equal to** (ℓ', v'), written as $(\ell, v) \succeq (\ell', v')$, if $(\ell, v) \succ (\ell', v')$ or $(\ell, v) = (\ell', v')$.*

*If $(\ell, v) \succ (\ell', v')$, then we also say that (ℓ', v') is **less than** (ℓ, v), written as $(\ell', v') \prec (\ell, v)$. The expression $(\ell', v') \preceq (\ell, v)$ is similarly defined.*

Lemma 4. *$\forall$ context $(n, V, \boldsymbol{\mathcal{V}}, \boldsymbol{\omega}) \in \mathscr{C}_n^V$, $\forall\ k \in \{1, \ldots, K+1\}$, $\forall$ player i, and $\forall$ strategy $s_i \in NSD_i^k(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)$, denoting s_i by (i, ℓ_i, v_i), we have*

$$(\ell_i, v_i) \succeq (\min\{\ell : g_i^\ell(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i) = g_i^{k-1}(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)\}, g_i^{k-1}(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)).$$

Proof. We prove Lemma 4 by induction on k. Arbitrarily fixing $(n, V, \boldsymbol{\mathcal{V}}, \boldsymbol{\omega})$, i, and $s_i \in NSD_i^k(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)$, we proceed by contradiction.

Assume $(\ell_i, v_i) \prec (\min\{\ell : g_i^\ell(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i) = g_i^{k-1}(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)\}, g_i^{k-1}(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i))$, that is, either $v_i < g_i^{k-1}(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)$, or $v_i = g_i^{k-1}(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)$ and $\ell_i > \min\{\ell : g_i^\ell(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i) = g_i^{k-1}(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)\}$. We shall prove that s_i is strongly dominated by another strategy $\hat{s}_i$ at $(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)$ with respect to $NSD_{-i}^{k-1}(\boldsymbol{\mathcal{V}}_i, \cdot)$, which contradicts the hypothesis $s_i \in NSD_i^k(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)$. The strategy $\hat{s}_i \triangleq (i, \hat{\ell}_i, \hat{v}_i)$ is such that

$$\hat{\ell}_i = \min\{\ell : g_i^\ell(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i) = g_i^{k-1}(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)\} \quad \text{and} \quad \hat{v}_i = g_i^{k-1}(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i).$$

Write $\boldsymbol{\mathcal{V}}_i = (\Omega, \mathbf{v}, \mathcal{P})$, and arbitrarily fix a state $\omega' \in \mathcal{P}_i(\boldsymbol{\omega}_i)$ and a strategy subprofile $t_{-i} \in NSD_{-i}^{k-1}(\boldsymbol{\mathcal{V}}_i, \omega')$. By definition it suffices to prove

$$u_i(\mathbf{v}(\omega'), (\hat{s}_i, t_{-i})) > u_i(\mathbf{v}(\omega'), (s_i, t_{-i})). \tag{7}$$

Let $\widehat{2^{nd}v}$ and $2^{nd}v$ respectively be the second-highest value announced by the players according to $(\hat{s}_i, t_{-i})$ and (s_i, t_{-i}). It is easy to see that $\widehat{2^{nd}v} \geq 2^{nd}v$, because each player's announced value either does not change or is stricter bigger, according to the first strategy profile. Let $\hat{\delta}_i$ and δ_i respectively be the rewards that player i gets in Step **c** in the plays of $(\hat{s}_i, t_{-i})$ and (s_i, t_{-i}). Because $\bar{\rho}$ is proper, we have

$$\hat{\delta}_i = \bar{\rho}(\widehat{2^{nd}v}, \hat{v}_i, \hat{\ell}_i) \geq \bar{\rho}(2^{nd}v, \hat{v}_i, \hat{\ell}_i) > \bar{\rho}(2^{nd}v, v_i, \ell_i) = \delta_i > 0.$$

Indeed, the first inequality above is because of Condition 1 in Definition 13, and the second is because of $(\ell_i, v_i) \prec (\hat{\ell}_i, \hat{v}_i)$ as well as Conditions 2 and 3 in Definition 13 —if $v_i < \hat{v}_i$ then the inequality holds by Condition 2, otherwise the inequality holds by Condition 3.

Let $(\hat{a}, \hat{P})$ and (a, P) respectively be the outcomes of the two plays, and denote t_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 of the induction ($k = 1$) and the Induction Step ($k > 1$). In this case we have $\hat{v}_i = g_i^{k-1}(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i) = g_i^0(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)$, and we distinguish three subcases.

Subcase 1.1. $a = i$.

In this subcase, we have $\hat{a} = i$ as well, because 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,

$$\begin{aligned} u_i(\mathbf{v}(\omega'), (\hat{s}_i, t_{-i})) &= \mathbf{v}_i(\omega') - \hat{P}_i = \mathbf{v}_i(\omega') - \max_{j \neq i} v'_j + \hat{\delta}_i \\ &> \mathbf{v}_i(\omega') - \max_{j \neq i} v'_j + \delta_i = \mathbf{v}_i(\omega') - P_i = u_i(\mathbf{v}(\omega'), (s_i, t_{-i})), \end{aligned}$$

where the inequality is because $\hat{\delta}_i > \delta_i$.

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

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

$$\begin{aligned} u_i(\mathbf{v}(\omega'), (\hat{s}_i, t_{-i})) &= \mathbf{v}_i(\omega') - \hat{P}_i = \mathbf{v}_i(\omega') - \max_{j \neq i} v'_j + \hat{\delta}_i \geq g_i^0(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i) - \max_{j \neq i} v'_j + \hat{\delta}_i \\ &= \hat{v}_i - \max_{j \neq i} v'_j + \hat{\delta}_i \geq \hat{\delta}_i > \delta_i = -P_i = u_i(\mathbf{v}(\omega'), (s_i, t_{-i})), \end{aligned}$$

where the first inequality is by the definition of $g_i^0(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)$, and the second is because $\hat{v}_i \geq \max_{j \neq i} v'_j$ as implied by the fact $\hat{a} = i$.

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

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

$$u_i(\mathbf{v}(\omega'), (\hat{s}_i, t_{-i})) = -\hat{P}_i = \hat{\delta}_i > \delta_i = -P_i = u_i(\mathbf{v}(\omega'), (s_i, t_{-i})).$$

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

This case applies to the Induction Step ($k > 1$) only. In this case, we shall prove that $\hat{a} \neq i$. To do so, first notice

$$g_i^{\hat{\ell}_i - 1}(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i) < g_i^{\hat{\ell}_i}(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i), \tag{8}$$

by the definition of $\hat{\ell}_i$. Second, notice

$$g_i^{k'}(\boldsymbol{\mathcal{V}}_i, \omega') = g_i^{k'}(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i) \ \forall \ k' \geq 0, \tag{9}$$

by Claim 1. Combining Equations 8 and 9 and taking $k' = \hat{\ell}_i - 1$, we have that

$$g_i^{\hat{\ell}_i - 1}(\boldsymbol{\mathcal{V}}_i, \omega') < g_i^{\hat{\ell}_i}(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i).$$

Because $g_i^{\hat{\ell}_i}(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i) = \min_{\omega'' \in \mathcal{P}_i(\boldsymbol{\omega}_i)} \max_{j'} g_{j'}^{\hat{\ell}_i - 1}(\boldsymbol{\mathcal{V}}_i, \omega'') \leq \max_{j'} g_{j'}^{\hat{\ell}_i - 1}(\boldsymbol{\mathcal{V}}_i, \omega')$, we have

$$g_i^{\hat{\ell}_i - 1}(\boldsymbol{\mathcal{V}}_i, \omega') < \max_{j'} g_{j'}^{\hat{\ell}_i - 1}(\boldsymbol{\mathcal{V}}_i, \omega').$$

Therefore, letting $j = \text{argmax}_{j'} g_{j'}^{\hat{\ell}_i - 1}(\boldsymbol{\mathcal{V}}_i, \omega')$ with ties broken lexicographically, we have

$$j \neq i \quad \text{and} \quad g_j^{\hat{\ell}_i - 1}(\boldsymbol{\mathcal{V}}_i, \omega') \geq g_i^{\hat{\ell}_i}(\boldsymbol{\mathcal{V}}_i, \omega),$$

and thus

$$(\hat{\ell}_i - 1, g_j^{\hat{\ell}_i - 1}(\boldsymbol{\mathcal{V}}_i, \omega')) \succ (\hat{\ell}_i, g_i^{\hat{\ell}_i}(\boldsymbol{\mathcal{V}}_i, \omega)). \tag{10}$$

Because $t_j \in NSD_j^{k-1}(\boldsymbol{\mathcal{V}}_i, \omega') \subseteq NSD_j^{\hat{\ell}_i}(\boldsymbol{\mathcal{V}}_i, \omega')$, by the induction hypothesis[13] we have

$$(\ell'_j, v'_j) \succeq (\min\{\ell : g_i^{\ell}(\boldsymbol{\mathcal{V}}_i, \omega') = g_i^{\hat{\ell}_i - 1}(\boldsymbol{\mathcal{V}}_i, \omega')\}, g_i^{\hat{\ell}_i - 1}(\boldsymbol{\mathcal{V}}_i, \omega')) \succeq (\hat{\ell}_i - 1, g_j^{\hat{\ell}_i - 1}(\boldsymbol{\mathcal{V}}_i, \omega')),$$

which together with Equation 10 implies that

$$(\ell'_j, v'_j) \succ (\hat{\ell}_i, g_i^{\hat{\ell}_i}(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)) = (\hat{\ell}_i, g_i^{k-1}(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)) = (\hat{\ell}_i, \hat{v}_i). \tag{11}$$

By Equation 11 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 $\hat{a} \neq i$. Since $(\hat{\ell}_i, \hat{v}_i) \succ (\ell_i, v_i)$, we have $a \neq i$ as well. Therefore in *Case 2*, we have $P_i = -\delta_i$ and $\hat{P}_i = -\hat{\delta}_i$, which implies

$$u_i(\mathbf{v}(\omega'), (\hat{s}_i, t_{-i})) = -\hat{P}_i = \hat{\delta}_i > \delta_i = -P_i = u_i(\mathbf{v}(\omega'), (s_i, t_{-i})).$$

In sum, Equation 7 holds, and s_i is strongly dominated by $\hat{s}_i$ at $(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)$ with respect to $NSD_{-i}^{k-1}(\boldsymbol{\mathcal{V}}_i, \cdot)$, contradicting the hypothesis that $s_i \in NSD_i^k(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)$. Therefore Lemma 4 holds. ■

Following Lemma 4, we have that for any order-$(k+1)$ rational strategy profile s in game $((\boldsymbol{\mathcal{V}}, \boldsymbol{\omega}), M)$, $2^{nd}v \geq G^k(\boldsymbol{\mathcal{V}}, \boldsymbol{\omega})$ and $rev(M(s)) = 2^{nd}v - \sum_i \delta_i$. Now the only remaining part in the proof of Lemma 3 is to show that M is possibilistic interim individually rational.

[13]The lemma is stated with respect to $(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)$, but it is easy to see that it can be stated with respect to any type profile $\mathcal{V}$ and state ω. Thus the induction hypothesis also applies to any type profile $\mathcal{V}$ and state ω —here $\boldsymbol{\mathcal{V}}_i$ and ω'.

Proof of Possibilistic Interim Individual Rationality of M. To do so, for each game framework $\mathcal{M} = (\Omega, \mathbf{v}, \mathcal{P}, \mathbf{s})$ for M, each player i, each state $\omega \in \Omega$, each state $\omega' \in \mathcal{P}_i(\omega)$, and each player $j \neq i$, denote $\mathbf{s}(\omega')_j$ by $(j, \ell_j(\omega'), v_j(\omega'))$.

If $\mathbf{v}(\omega')_i \geq \max_{j \neq i} v_j(\omega')$ for each $\omega' \in \mathcal{P}_i(\omega)$, then $\mathbf{safe}_i(\omega) \triangleq (i, V, 0)$. By doing so, player i gets the good in each state he considers possible, pays no more than his valuation in that state, and maximizes the reward he gets in Step **c** of the mechanism, which is always positive. It is easy to verify that $\mathbf{safe}_i(\omega)$ satisfies the two conditions in Definition 9.

If there exists $\omega' \in \mathcal{P}_i(\omega)$ such that $\mathbf{v}(\omega')_i < \max_{j \neq i} v_j(\omega')$, then for each such ω', let $j(\omega')$ be the player whose announced triple at state ω' is ordered the first by M among $-i$. That is, $j(\omega') = \text{argmax}_{j' \neq i} v_{j'}(\omega')$, with ties broken in favor of smaller value of $\ell_{j'}(\omega')$, and further in favor of smaller player identity. Let strategy $s_i(\omega') = (i, v_i(\omega'), \ell_i(\omega'))$ be such that, it maximizes player i's reward subject to the following condition: when player i uses $s_i(\omega')$ and each $j \neq i$ uses $\mathbf{s}(\omega')_j$, player $j(\omega')$'s announced triple is ordered the first by M among all players. Let $\mathbf{safe}_i(\omega)$ be the strategy that is ordered the last by M, in the set

$$\hat{S} \triangleq \{s_i(\omega') : \omega' \in \mathcal{P}_i(\omega), \mathbf{v}(\omega')_i < \max_{j \neq i} v_j(\omega')\}.$$

It is easy to verify that $\mathbf{safe}_i(\omega)$ satisfies Condition 2 of Definition 9, i.e., player i believes that his utility is always non-negative by using $\mathbf{safe}_i(\omega)$. Indeed, at any state $\omega' \in \mathcal{P}_i(\omega)$ player i always gets non-negative reward, and he wins the good only if $\omega' \notin \hat{S}$, in which case his price is $\max_{j \neq i} v_j(\omega') \leq \mathbf{v}(\omega')_i$. As for Condition 1 of Definition 9, notice that if $\mathbf{safe}_i(\omega)$ is not strongly dominated at ω with respect to S'_{-i} where $S'_{-i}(\hat{\omega}) = \{\mathbf{s}(\hat{\omega})_{-i}\}$ for each $\hat{\omega} \in \Omega$, then $\mathbf{safe}_i(\omega)$ satisfies Condition 1. Otherwise, there exists a pure strategy $\hat{s}_i$ strongly dominating $\mathbf{safe}_i(\omega)$ at ω with respect to S'_{-i}. Resetting $\mathbf{safe}_i(\omega)$ to be $\hat{s}_i$, we have that player i's utility at each $\omega' \in \mathcal{P}_i(\omega)$ by using $\mathbf{safe}_i(\omega)$ has increased, and thus $\mathbf{safe}_i(\omega)$ still satisfies Condition 2. Repeatedly check whether $\mathbf{safe}_i(\omega)$ is strongly dominated at ω with respect to S'_{-i} and replace it with the dominating strategy if so. Because there are finitely many pure strategies, ultimately we shall find a $\mathbf{safe}_i(\omega)$ that is not strongly dominated, and this strategy satisfies both conditions in Definition 9.

Accordingly, M is possibilistic interim individually rational, concluding the proof of Lemma 3. ■

6.5 Proofs of Theorems 2 and 3

Let us now argue that the Arbitrary-Belief Lemma indeed implies Theorems 2 and 3.

Proof of Theorem 2. Because the reward function ρ is proper, the Arbitrary-Belief Lemma implies that the mechanism $M_{n,V,K,\varepsilon}$ is possibilistic interim individually rational.

By Theorem 1, for any $k \in \{0, \ldots, K\}$, to show that $M_{n,V,K,\varepsilon}$ order-$(k+1)$ rationally implements $G^k - \varepsilon$, it suffices to show that $\forall$ context $(n, V, \boldsymbol{\mathcal{V}}, \boldsymbol{\omega}) \in \mathscr{C}_n^V$ and $\forall$ strategy profile $s \in \prod_i NSD_i^{k+1}(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i)$, we have

$$rev(M_{n,V,K,\varepsilon}(s)) \geq G^k(\boldsymbol{\mathcal{V}}, \boldsymbol{\omega}) - \varepsilon.$$

Denote s_i by (i, ℓ_i, v_i) for each player i, the second highest value announced by the players in the play of s by $2^{nd}v$, the reward of each player i in Step **c** by δ_i, and the final

outcome $M_{n,V,K,\varepsilon}(s)$ by (a, P). Again by Lemma 3 we have that $2^{nd}v \geq G^k(\boldsymbol{\mathcal{V}}, \boldsymbol{\omega})$ and $rev(M_{n,V,K,\varepsilon}(s)) = 2^{nd}v - \sum_i \delta_i$. Because for each player i

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

we have

$$rev(M_{n,V,K,\varepsilon}(s)) = 2^{nd}v - \sum_i \delta_i \geq G^k(\boldsymbol{\mathcal{V}}, \boldsymbol{\omega}) - \sum_i \delta_i \geq G^k(\boldsymbol{\mathcal{V}}, \boldsymbol{\omega}) - \sum_i \frac{\varepsilon}{n} = G^k(\boldsymbol{\mathcal{V}}, \boldsymbol{\omega}) - \varepsilon.$$

Thus Theorem 2 holds. ■

Proof of Theorem 3. Consider the following reward function:

$$\rho'(2^{nd}v, v_i, \ell_i) \triangleq \frac{\varepsilon(2^{nd}v + 1)}{2n}\left[1 + \frac{v_i}{1+v_i} - \frac{\ell_i}{(1+\ell_i)(1+V)^2}\right],$$

and let $M'_{n,V,K,\varepsilon}$ be the mechanism obtained by replacing ρ with ρ' in $M_{n,V,K,\varepsilon}$. Theorem 3 then easily follows from the fact that ρ' is proper and that $\delta_i \leq \frac{\varepsilon(2^{nd}v+1)}{n}$ for each i. ■

Additional Ways to Trade Revenue for Robustness Using different proper reward function within the Arbitrary-Belief Mechanism gives the seller much flexibility in trading revenue for "player motivation." In particular, he can use the reward function ρ' just for the winner and the second-highest bidder, and use ρ for all others.

7 Impossibility Results for Epistemic Implementation

Theorem 4. *For any n, V, k and $c < V$, there is no possibilistic interim individually rational mechanism that order-k rationally implements $G^k - c$ for single-good auctions with n players and valuation bound 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 possibilistic interim individually rational mechanism M. It suffices to show that there exists an auction context $C = (2, V, \boldsymbol{\mathcal{V}}, \boldsymbol{\omega})$ such that in game (C, M) the following statement holds:

$$\exists \textit{strategy profile } s \in \textstyle\prod_i NSD_i^k(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i) \textit{ such that } rev(M(s)) < G^k(\boldsymbol{\mathcal{V}}, \boldsymbol{\omega}) - c. \tag{12}$$

To construct C, we construct a valuation framework $\mathcal{V} = (\Omega, \mathbf{v}, \mathcal{P})$ as follows.

- $\Omega = \{\omega\} \cup \{\omega_{i,\ell} : i \in \{1, 2\}, \ell \in \{1, 2, \ldots, k\}\}$;
- $\mathbf{v}(\omega) = (0, 0), \mathbf{v}(\omega_{i,\ell}) = (0, 0)\ \forall i$ and $\forall \ell < k$,
  if k is odd then $\mathbf{v}(\omega_{1,k}) = (0, V)$ and $\mathbf{v}(\omega_{2,k}) = (V, 0)$,
  otherwise $\mathbf{v}(\omega_{1,k}) = (V, 0)$ and $\mathbf{v}(\omega_{2,k}) = (0, V)$;

- $\mathcal{P}_1(\omega) = \{\omega_{1,1}\}$,
  $\mathcal{P}_1(\omega_{1,\ell}) = \{\omega_{1,\ell}\}$ $\forall$ odd $\ell < k$, $\mathcal{P}_1(\omega_{1,\ell}) = \{\omega_{1,\ell+1}\}$ $\forall$ even $\ell < k$, $\mathcal{P}_1(\omega_{1,k}) = \{\omega_{1,k}\}$,
  $\mathcal{P}_1(\omega_{2,\ell}) = \{\omega_{2,\ell+1}\}$ $\forall$ odd $\ell < k$, $\mathcal{P}_1(\omega_{2,\ell}) = \{\omega_{2,\ell}\}$ $\forall$ even $\ell < k$, and $\mathcal{P}_1(\omega_{2,k}) = \{\omega_{2,k}\}$;
- $\mathcal{P}_2(\omega) = \{\omega_{2,1}\}$,
  $\mathcal{P}_2(\omega_{1,\ell}) = \{\omega_{1,\ell+1}\}$ $\forall$ odd $\ell < k$, $\mathcal{P}_2(\omega_{1,\ell}) = \{\omega_{1,\ell}\}$ $\forall$ even $\ell < k$, $\mathcal{P}_2(\omega_{1,k}) = \{\omega_{1,k}\}$,
  $\mathcal{P}_2(\omega_{2,\ell}) = \{\omega_{2,\ell}\}$ $\forall$ odd $\ell < k$, $\mathcal{P}_2(\omega_{2,\ell}) = \{\omega_{2,\ell+1}\}$ $\forall$ even $\ell < k$, $\mathcal{P}_2(\omega_{2,k}) = \{\omega_{2,k}\}$.

We let $\boldsymbol{\mathcal{V}}_1 = \boldsymbol{\mathcal{V}}_2 = \mathcal{V}$ and $\boldsymbol{\omega}_1 = \boldsymbol{\omega}_2 = \omega$.

When k is odd, the valuation framework is illustrated by Figure 3, which follows our graphical representation except that some self-loops are omitted for succinctness —any missing edge at any node corresponds to a self-loop at that node.

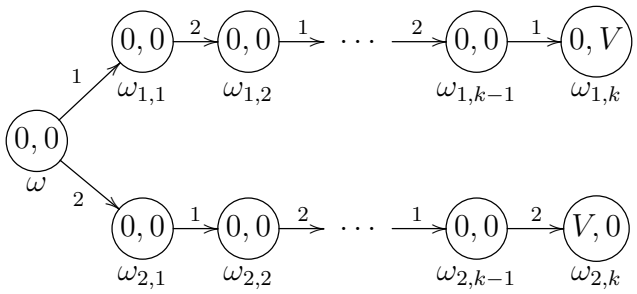


Figure 3: Valuation framework $\mathcal{V}$ when k is odd.

To prove Statement 12, we introduce an auxiliary context $\widehat{C} = (n, V, \widehat{\boldsymbol{\mathcal{V}}}, \boldsymbol{\omega})$, where:

- $\widehat{\boldsymbol{\mathcal{V}}}_1 = \widehat{\boldsymbol{\mathcal{V}}}_2$ and both equal an auxiliary valuation framework $\widehat{\mathcal{V}} = (\Omega, \hat{\mathbf{v}}, \mathcal{P})$, such that $\hat{\mathbf{v}}$ equals $\mathbf{v}$ everywhere except $\hat{\mathbf{v}}(\omega_{1,k}) = \hat{\mathbf{v}}(\omega_{2,k}) = (0, 0)$; and
- $\boldsymbol{\omega}$ is the same one as in C.

When k is odd, $\widehat{\mathcal{V}}$ is shown in Figure 4.

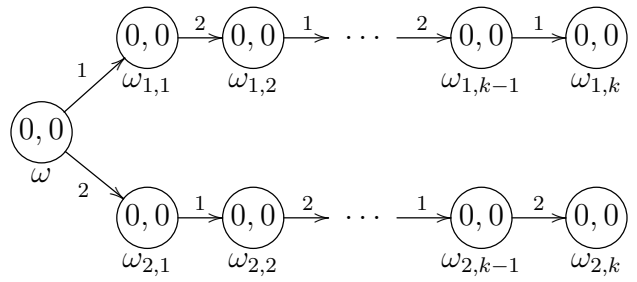


Figure 4: Valuation framework $\widehat{\mathcal{V}}$ when k is odd.

By definition we have that for each player i, $g_i^k(\mathcal{V}, \omega) = V$ and $g_i^k(\widehat{\mathcal{V}}, \omega) = 0$. Thus

$$G^k(\boldsymbol{\mathcal{V}}, \boldsymbol{\omega}) = V \quad \text{and} \quad G^k(\widehat{\boldsymbol{\mathcal{V}}}, \boldsymbol{\omega}) = 0.$$

However, we shall show that the order-k rational strategies are the same at both $(\mathcal{V}, \omega)$ and $(\widehat{\mathcal{V}}, \omega)$. To do so, note that for each $\omega' \in \Omega$,

$$NSD^0(\mathcal{V}, \omega') = NSD^0(\widehat{\mathcal{V}}, \omega') = S,$$

where S is the strategy space of M. Because $\mathbf{v}(\omega')_i = \hat{\mathbf{v}}(\omega')_i = 0$ for every player i and every state ω' except $\omega_{1,k}$ and $\omega_{2,k}$, by the definition of the iterated deletion procedure and the construction of $\mathcal{P}_1$ and $\mathcal{P}_2$, we have

$$NSD^1(\mathcal{V}, \omega') = NSD^1(\widehat{\mathcal{V}}, \omega') \ \forall \omega' \in \{\omega\} \cup \{\omega_{i,\ell} : i \in \{1,2\}, \ell \leq k-1\}.$$

By induction, we further have that for each $\ell' < k$,

$$NSD^{\ell'}(\mathcal{V}, \omega') = NSD^{\ell'}(\widehat{\mathcal{V}}, \omega') \ \forall \omega' \in \{\omega\} \cup \{\omega_{i,\ell} : i \in \{1,2\}, \ell \leq k-\ell'\}.$$

In particular, $NSD^{k-1}(\mathcal{V}, \omega) = NSD^{k-1}(\widehat{\mathcal{V}}, \omega)$ and $NSD^{k-1}(\mathcal{V}, \omega_{i,1}) = NSD^{k-1}(\widehat{\mathcal{V}}, \omega_{i,1})$ for each i. Finally, this implies that

$$NSD^k(\mathcal{V}, \omega) = NSD^k(\widehat{\mathcal{V}}, \omega).$$

Following this equation, and because $G^k(\mathcal{V}, \omega) - c = V - c > 0$, to prove Statement 12 it suffices to show that there exists a strategy profile $s \in NSD^k(\widehat{\mathcal{V}}, \omega)$ such that $rev(M(s)) \leq 0$. Since $\hat{\mathbf{v}}(\omega) = (0,0)$, we have $rev(M(s)) = -u_1(\hat{\mathbf{v}}(\omega), s) - u_2(\hat{\mathbf{v}}(\omega), s)$. Therefore it suffices to show the following statement:

$$\exists s \in NSD^k(\widehat{\mathcal{V}}, \omega) \textit{ such that } u_i(\hat{\mathbf{v}}(\omega), s) \geq 0 \textit{ for each } i. \tag{13}$$

To do so, notice that it is easy to construct a game framework $\mathcal{M} = (\Omega', \mathbf{v}', \mathcal{P}', \mathbf{s})$ consistent with $\widehat{\mathcal{V}}$ under some consistency mapping ψ, satisfying the following requirement: there exists $\omega' \in \Omega'$ such that

- $\psi(\omega') = \omega$; and
- $\mathbf{s}(\mathcal{P}'_i(\omega'))_{-i} = S_{-i}$ for each player i —that is, player i believes that the other player may use any strategy.

Since M is possibilistic interim individually rational, for each i there exists strategy $\mathbf{safe}_i(\omega')$ such that

$$\forall s'_i, \exists \omega'' \in \mathcal{P}'_i(\omega') \text{ such that } u_i(\mathbf{v}'(\omega''), (\mathbf{safe}_i(\omega'), \mathbf{s}(\omega'')_{-i})) \geq u_i(\mathbf{v}'(\omega''), (s'_i, \mathbf{s}(\omega'')_{-i})) \tag{14}$$

and

$$\forall \omega'' \in \mathcal{P}'_i(\omega'), u_i(\mathbf{v}'(\omega''), (\mathbf{safe}_i(\omega'), \mathbf{s}(\omega'')_{-i})) \geq 0. \tag{15}$$

Because $\psi(\mathcal{P}'_i(\omega')) = \mathcal{P}_i(\psi(\omega')) = \mathcal{P}_i(\omega) = \{\omega_{i,1}\}$, we have $\mathbf{v}'(\omega'') = \hat{\mathbf{v}}(\omega_{i,1}) = (0,0)$ for each $\omega'' \in \mathcal{P}'_i(\omega')$. Thus Equation 14 implies that $\mathbf{safe}_i(\omega')$ is not strongly dominated at $(\widehat{\mathcal{V}}, \omega)$ with respect to $NSD^0_{-i}(\widehat{\mathcal{V}}, \cdot)$, that is,

$$\mathbf{safe}_i(\omega') \in NSD^1_i(\widehat{\mathcal{V}}, \omega).$$

Further, because $\mathbf{s}(\mathcal{P}'_i(\omega'))_{-i} = S_{-i} = NSD^0_{-i}(\widehat{\mathcal{V}}, \omega_{i,1})$, Equation 15 implies

$$u_i(\hat{\mathbf{v}}(\omega_{i,1}), (\mathbf{safe}_i(\omega'), s_{-i})) \geq 0 \ \forall s_{-i} \in NSD^0_{-i}(\widehat{\mathcal{V}}, \omega_{i,1}).$$

Because $NSD^1_{-i}(\widehat{\mathcal{V}}, \omega_{i,1}) \subseteq NSD^0_{-i}(\widehat{\mathcal{V}}, \omega_{i,1})$, we have

$$u_i(\hat{\mathbf{v}}(\omega_{i,1}), (\mathbf{safe}_i(\omega'), s_{-i})) \geq 0 \ \forall s_{-i} \in NSD^1_{-i}(\widehat{\mathcal{V}}, \omega_{i,1}).$$

Below we consider a new strategy $\hat{s}_i$. If $\mathbf{safe}_i(\omega') \in NSD^2_i(\widehat{\mathcal{V}}, \omega)$ then let $\hat{s}_i = \mathbf{safe}_i(\omega')$. Otherwise let $\hat{s}_i$ be an arbitrary strategy in $NSD^2_i(\widehat{\mathcal{V}}, \omega)$ strongly dominating it at $(\widehat{\mathcal{V}}, \omega)$ with respect to $NSD^1_{-i}(\widehat{\mathcal{V}}, \cdot)$. By construction we always have

$$\hat{s}_i \in NSD^2_i(\widehat{\mathcal{V}}, \omega) \quad \text{and} \quad u_i(\hat{\mathbf{v}}(\omega_{i,1}), (\hat{s}_i, s_{-i})) \geq 0 \ \forall s_{-i} \in NSD^1_{-i}(\widehat{\mathcal{V}}, \omega_{i,1}).$$

Continuing this line of reasoning, we finally have that for each i there exists strategy s_i such that

$$\hat{s}_i \in NSD^k_i(\widehat{\mathcal{V}}, \omega) \quad \text{and} \quad u_i(\hat{\mathbf{v}}(\omega_{i,1}), (\hat{s}_i, s_{-i})) \geq 0 \ \forall s_{-i} \in NSD^{k-1}_{-i}(\widehat{\mathcal{V}}, \omega_{i,1}).$$

Accordingly, the strategy profile $\hat{s} = (\hat{s}_1, \hat{s}_2)$ is in $NSD^k(\widehat{\mathcal{V}}, \omega)$. Because $\hat{\mathbf{v}}(\omega) = \hat{\mathbf{v}}(\omega_{i,1}) = (0, 0)$, to prove Statement 13, the only thing remaining to show is

$$\hat{s}_{-i} \in NSD^{k-1}_{-i}(\widehat{\mathcal{V}}, \omega_{i,1}) \ \forall i.$$

Because $\hat{s}_{-i} \in NSD^k_{-i}(\widehat{\mathcal{V}}, \omega) \subseteq NSD^{k-1}_{-i}(\widehat{\mathcal{V}}, \omega)$, to prove Statement 13 it suffices to show that for each i,

$$NSD^{k-1}_{-i}(\widehat{\mathcal{V}}, \omega) = NSD^{k-1}_{-i}(\widehat{\mathcal{V}}, \omega_{i,1}).$$

This follows from the fact that $\hat{\mathbf{v}}$ maps every state to the valuation profile $(0, 0)$, and thus $NSD^{k-1}_{-i}(\widehat{\mathcal{V}}, \cdot)$ is the same at every state. Therefore Statement 13 holds, and so does Statement 12. Accordingly, Theorem 4 holds for $n = 2$ and $k > 0$.

In the degenerated case where $k = 0$, using the same idea as before, consider contexts $C = (\boldsymbol{\mathcal{V}}, \boldsymbol{\omega})$ and $\widehat{C} = (\widehat{\boldsymbol{\mathcal{V}}}, \boldsymbol{\omega})$. let $\boldsymbol{\mathcal{V}}_1 = \boldsymbol{\mathcal{V}}_2 = \mathcal{V} = (\Omega, \mathbf{v}, \mathcal{P})$ be such that $\Omega = \{\omega\}$ and $\mathbf{v}(\omega) = (V, V)$, and $\widehat{\boldsymbol{\mathcal{V}}}_1 = \widehat{\boldsymbol{\mathcal{V}}}_2 = \widehat{\mathcal{V}} = (\Omega, \hat{\mathbf{v}}, \mathcal{P})$ be such that $\hat{\mathbf{v}}(\omega) = (0, 0)$. Because M is possibilistic interim individually rational, in game $(\widehat{C}, M)$ there exists a strategy profile s such that $u_i(\hat{\mathbf{v}}(\omega), s) \geq 0$ for each i. But then $rev(M(s)) \leq 0 < V - c = G^0(\boldsymbol{\mathcal{V}}, \boldsymbol{\omega}) - c$. Because $s \in \prod_i NSD^0_i(\boldsymbol{\mathcal{V}}_i, \boldsymbol{\omega}_i) = S$, M cannot order-0 rationally implement $G^0 - c$.

In sum, Theorem 4 holds for $n = 2$. For $n > 2$, we construct the desired contexts and valuation frameworks by adding dummy players to the valuation frameworks $\mathcal{V}$ and $\widehat{\mathcal{V}}$ of the 2-player case. The analysis is almost the same, and thus omitted. ■

8 Concluding Remarks

In this paper we have extended the notions of implementation and social choice correspondence, so as to incorporate the players' arbitrary higher-order beliefs about (payoff) types. In so doing we hope to have established a tighter connection between implementation and epistemic game theory. Indeed there are plenty of attractive epistemic social choice correspondences whose implementability should be investigated.

Also, we can further extend the notion of an epistemic social correspondence by allowing it to additionally depend on the players' maximum common rationality order. Indeed, it

is trivial to (formally state and) verify that the Arbitrary-Belief Mechanism actually implements the revenue benchmark $\mathcal{G} \triangleq G^m$, where m is the maximum integer k such that all players are order-k rational, although no one (player or designer) might know the actual value of m.

Acknowledgements

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.

References

[1] D. Abreu and H. Matsushima. Virtual Implementation in Iteratively Undominated Strategies: Complete Information. *Econometrica*, Vol. 60, No. 5, pp. 993-1008, 1992.

[2] M. Allais. Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine. *Econometrica*, Vol. 21, No. 4, pp. 503-546, 1953.

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

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

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

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

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

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

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

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

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

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

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

[14] J. Glazer and M. Perry. Virtual Implementation in Backwards Induction. *Games and Economic Behavior*, Vol.15, pp. 27-32, 1996.

[15] J. Halpern and R. Pass. A Logical Characterization of Iterated Admissibility. *Conference on Theoretical Aspects of Rationality and Knowledge (TARK)*, pp. 146-155, 2009.

[16] J. Harsanyi. Games with Incomplete Information Played by "Bayesian" Players, I-III. Part I. The Basic Model. *Management Science*, 14(3) Theory Series: 159-182, 1967.

[17] M. Jackson. Implementation in Undominated Strategies: A Look at Bounded Mechanisms. *The Review of Economic Studies*, 59(4): 757-775, 1992.

[18] 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.

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

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

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

[22] R. Wilson. Game-theoretic analyses of trading processes. *In T. F. Bewley (ed.), Advances in Economic Theory, Fifth World Congress*, Cambridge University Press, Cambridge, UK, pp. 33-70, 1987.