Supplement to "Tight Revenue Bounds With Possibilistic Beliefs and Level-k Rationality"*

Jing Chen† Silvio Micali‡ Rafael Pass§

S1 Proof of Theorem 1

We break our proof into simpler claims.

Claim 1. *M is IIR.*

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

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

In the outcome of (a_i, a'_{-i}), if $w \neq i$, then $P_i = -\delta_i$, and thus $u_i(a_i, a'_{-i}) = -P_i = \delta_i > 0$. If $w = i$, then $\theta_i \geq 2^{nd}v$ and $P_i = 2^{nd}v - \delta_i$, thus

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

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

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

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

if $v > v'$ or ($v = v'$ and $\ell < \ell'$). We write $(\ell, v) \succeq (\ell', v')$ if $(\ell, v) \succ (\ell', v')$ or $(\ell, v) = (\ell', v')$. Notice that the relation defined by "$\succ$" is complete: for any two pairs $(\ell, v) \neq (\ell', v')$, either $(\ell, v) \succ (\ell', v')$ or $(\ell', v') \succ (\ell, v)$. Also notice that the order defined by "$\succ$" is consistent with how the mechanism breaks ties.

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

*The authors thank Shafi Goldwasser, Aviad Heifetz, Andrew Lo, Ron Rivest, the editor of Econometrica and several wonderful anonymous referees for many helpful comments, and in particular one referee for discussions now reported in Section 5.2. The third author thanks Joseph Halpern for introducing him to the area of epistemic game theory, and for hours and hours of enlightening discussions about it. The first two authors have been partially supported by ONR Grant No. N00014-09-1-0597. The third author has been partially supported by an Alfred P. Sloan Fellowship, Microsoft New Faculty Fellowship, NSF Award CNS-1217821, NSF CAREER Award CCF-0746990, NSF Award CCF-1214844, AFOSR YIP Award FA9550-10-1-0093, and DARPA and AFRL under contract FA8750-11-2-0211.

†Department of Computer Science, Stony Brook University, Stony Brook, NY 11794, USA. jingchen@cs.stonybrook.edu.

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

§Department of Computer Science, Cornell University, Ithaca, NY 14853, USA. rafael@cs.cornell.edu.

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

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

If $v_i > v_i'$, we have

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

where the first inequality holds because $0 \leq v_i' < v_i$ and $\ell_i' \geq 0$, the second because $0 \leq v_i' < v_i$ and both v_i and v_i' are integers, and the third because $\frac{\ell_i' - \ell_i}{(1+\ell_i)(1+\ell_i')} \geq \frac{-\ell_i}{(1+\ell_i)(1+\ell_i')} \geq \frac{-\ell_i}{1+\ell_i} > -1$ and $1 + v_i > 0$. Thus $\delta_i > \delta_i'$ as desired.

If $v_i = v_i'$ and $\ell_i < \ell_i'$, we have

$$\delta_i - \delta_i' = \frac{\varepsilon}{2n} \cdot \frac{\ell_i' - \ell_i}{(1+\ell_i)(1+\ell_i')(1+v_i)^2} > 0,$$

thus again $\delta_i > \delta_i'$.

Therefore Claim 2 holds. □

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

Claim 3. $\forall\, k \in \{1, \ldots, K+1\}$ *we have that*

$$\forall a_i = (i, \ell_i, v_i) \in RAT_i^k(\tau_i),\ (\ell_i, v_i) \succeq (\min\{\ell : g_i^\ell(\tau_i) = g_i^{k-1}(\tau_i)\}, g_i^{k-1}(\tau_i)). \tag{2}$$

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

Assume Equation 2 holds for all $k' < k$. To prove it for k, we proceed by contradiction. Let $\hat{\ell}_i = \min\{\ell : g_i^\ell(\tau_i) = g_i^{k-1}(\tau_i)\}$ and $\hat{v}_i = g_i^{k-1}(\tau_i)$, and assume $(\ell_i, v_i) \not\succeq (\hat{\ell}_i, \hat{v}_i)$. By the definition of "$\succ$" we have $(\hat{\ell}_i, \hat{v}_i) \succ (\ell_i, v_i)$.

Let $\hat{a}_i \triangleq (i, \hat{\ell}_i, \hat{v}_i)$, and arbitrarily fix $t_{-i} \in B_i(\tau_i)$ and $a'_{-i} \in RAT_{-i}^{k-1}(t_{-i})$. Below we show

$$u_i((\hat{a}_i, a'_{-i}), \theta_i) > u_i((a_i, a'_{-i}), \theta_i), \tag{3}$$

which contradicts the fact $a_i \in RAT_i^k(\tau_i)$.

To prove Equation 3, let $\hat{\delta}_i$ and δ_i respectively be the rewards that player i gets in Step **c** of the mechanism according to $(\hat{a}_i, a'_{-i})$ and (a_i, a'_{-i}). Because $(\hat{\ell}_i, \hat{v}_i) \succ (\ell_i, v_i)$, by Claim 2 we have

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

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

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

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

Subcase 1.1. $w = i$.

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

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

where the inequality holds because $\hat{\delta}_i > \delta_i$. Thus Equation 3 holds.

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

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

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

and Equation 3 holds.

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

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

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

and again Equation 3 holds.

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

This case applies to the Induction Step only. (In the Base Case, we have $\hat{\ell}_i = 0$.)
In this case, we shall prove that $\hat{w} \neq i$. To do so, first note that, by the definition of $\hat{\ell}_i$,

$$g_i^{\hat{\ell}_i - 1}(\tau_i) < g_i^{\hat{\ell}_i}(\tau_i). \tag{4}$$

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

$$g_i^{\hat{\ell}_i}(\tau_i) = \min_{t'_{-i} \in B_i(\tau_i)} \max \left\{ \left(g_i^{\hat{\ell}_i - 1}(\tau_i), g_{-i}^{\hat{\ell}_i - 1}(t'_{-i}) \right) \right\} \leq \max \left\{ \left(g_i^{\hat{\ell}_i - 1}(\tau_i), g_{-i}^{\hat{\ell}_i - 1}(t_{-i}) \right) \right\}. \tag{5}$$

Combining Equations 4 and 5, we have

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

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

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

Accordingly,

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

thus

$$(\hat{\ell}_i - 1, g_j^{\hat{\ell}_i - 1}(t_j)) \succ (\hat{\ell}_i, g_i^{\hat{\ell}_i}(\tau_i)). \tag{6}$$

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

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

which together with Equation 6 implies

$$(\ell'_j, v'_j) \succ (\hat{\ell}_i, g_i^{\hat{\ell}_i}(\tau_i)) = (\hat{\ell}_i, g_i^{k-1}(\tau_i)) = (\hat{\ell}_i, \hat{v}_i). \tag{7}$$

By Equation 7 we have that the triple (j, ℓ'_j, v'_j) is ordered before $(i, \hat{\ell}_i, \hat{v}_i)$ according to M, and thus $\hat{w} \neq i$. Since $(\hat{\ell}_i, \hat{v}_i) \succ (\ell_i, v_i)$, we have $w \neq i$ as well. Therefore $P_i = -\delta_i$ and $\hat{P}_i = -\hat{\delta}_i$, which implies

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

Thus Equation 3 holds.

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

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

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

we have

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

This concludes the proof of Theorem 1. ■

S2 Proof of Theorem 2

We first prove the theorem for $n = 2$. Arbitrarily fix $V, k \geq 1$ (the case where $k = 0$ is degenerated and will be briefly discussed at the end) and $c < V$. Assuming there exists an IIR mechanism $\hat{M}$ that level-k rationally implements $G^k - c$ for $\mathscr{C}_{n,V}$, we prove the following statement:

$$\textit{There exist } C = (2, V, \mathcal{T}, \tau) \in \mathscr{C}_{n,V} \textit{ and } a \in RAT^k(\tau) \textit{ s.t. } rev(\hat{M}(a)) < G^k(C) - c, \tag{8}$$

which leads to a contradiction. To prove Statement 8, we set $\mathcal{T} = (T, \Theta, \nu, B)$ as follows: for each player i,

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

The type structure $\mathcal{T}$ is illustrated in Figure 1, and we set $\tau_i = t_{i,0}$ for each i.

Below, we show that there exists an action profile $a \in RAT^k(\tau)$ such that $rev(\hat{M}(a)) < G^k(C) - c$. For doing so, we use an auxiliary context $C' = (2, V, \mathcal{T}', \tau')$, where $\mathcal{T}' = (T', \Theta, \nu', B')$ is defined as follows: for each player i,

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

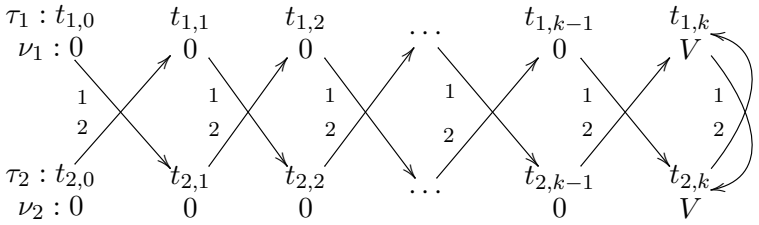


Figure 1: Type structure $\mathcal{T}$ in context C.

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

The type structure $\mathcal{T}'$ is illustrated in Figure 2, and we set $\tau'_i = t'_{i,0}$ for each i.

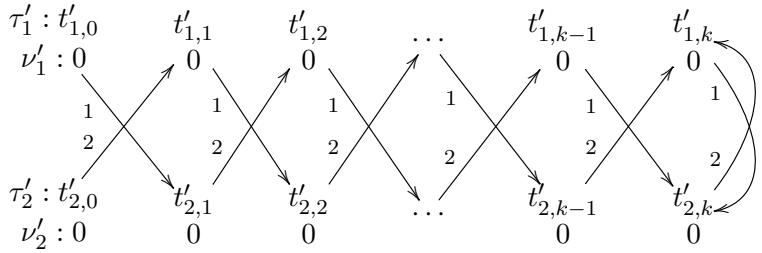


Figure 2: Type structure $\mathcal{T}'$ in context C'.

We first prove the following claim.

Claim 4. *In type structure $\mathcal{T}$, for any player i and any $\ell, k' \in \{0, 1, \ldots, k\}$, $g_i^{k'}(t_{i,\ell}) = 0$ if $k' + \ell < k$ and $g_i^{k'}(t_{i,\ell}) = V$ otherwise.*

Proof. We proceed by an induction on k'. The case with $k' = 0$ holds immediately, since $g_i^0(t_{i,\ell}) = \nu_i(t_{i,\ell})$, which is 0 when $\ell < k$ and V otherwise. For $k' \geq 1$, assuming the case is true for $k' - 1$, we show that it is true for k' as well. Indeed, for any player i,

$$g_i^{k'}(t_{i,k}) = \max\{g_i^{k'-1}(t_{i,k}), g_{-i}^{k'-1}(t_{-i,k})\} = \max\{V, V\} = V,$$

where the second equality is by the inductive hypothesis and the fact that $k' - 1 + k \geq k$. For any $\ell < k$, we have $g_i^{k'}(t_{i,\ell}) = \max\{g_i^{k'-1}(t_{i,\ell}), g_{-i}^{k'-1}(t_{-i,\ell+1})\}$. If $k' + \ell < k$, then $(k' - 1) + \ell < k$ and $(k' - 1) + (\ell + 1) = k' + \ell < k$; thus, by the inductive hypothesis, we have

$$g_i^{k'}(t_{i,\ell}) = \max\{0, 0\} = 0.$$

If $k' + \ell \geq k$, then $(k' - 1) + (\ell + 1) \geq k$; thus, by the inductive hypothesis, we have

$$g_i^{k'}(t_{i,\ell}) = \max\{g_i^{k'-1}(t_{i,\ell}), V\} = V,$$

where the second equality is because $g_i^{k'-1}(t_{i,\ell}) \leq V$. Therefore Claim 4 holds. □

By Claim 4, $g_i^k(t_{i,0}) = V$ for each i; thus,

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

Accordingly, to prove Statement 8 it suffices to prove the following two propositions:

$$RAT^k(\tau) = RAT^k(\tau'), \tag{9}$$

and

$$\text{there exists } a \in RAT^k(\tau') \text{ such that } rev(\hat{M}(a)) \leq 0. \tag{10}$$

To prove Equation 9, recall that, by definition,

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

where A_i is the set of actions for player i in $\hat{M}$. Because $\nu_i(t_{i,\ell}) = \nu'_i(t'_{i,\ell}) = 0$ for each i and each $\ell < k$, and because of the definitions of B and B', by a similar induction as the one in the proof of Claim 4 we have that, for any player i and any $\ell, k' \in \{0, 1, \ldots, k\}$,

$$RAT_i^{k'}(t_{i,\ell}) = RAT_i^{k'}(t'_{i,\ell}) \text{ whenever } k' + \ell \leq k.$$

In particular, $RAT_i^k(t_{i,0}) = RAT_i^k(t'_{i,0})$ for each i, and Equation 9 holds.

To prove Statement 10, notice that $\nu'_i(\tau'_i) = 0$ for each player i. Thus, for each action profile a, we have $rev(\hat{M}(a)) = -u_1(a, 0) - u_2(a, 0)$. Accordingly, it suffices to prove the following statement:

$$\text{there exists } a \in RAT^k(\tau') \text{ such that } u_i(a, 0) \geq 0 \text{ for each } i. \tag{11}$$

Since $\hat{M}$ is IIR, for each player i, there exists an action a_i such that

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

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

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

Indeed, if $a_i \in RAT_i^1(\tau'_i)$, then $a_i^1 = a_i$; else a_i^1 is the action in $RAT_i^1(\tau'_i)$ that dominates a_i.

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

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

Note that $a^k \in RAT^k(\tau')$. Accordingly, to prove Statement 11 it suffices to show that $a_{-i}^k \in RAT_{-i}^{k-1}(t'_{-i,1})$ for each i, because then we have $u_i(a^k, 0) \geq 0$ for each i, as desired. Thus it is left to show

$$a_i^k \in RAT_i^{k-1}(t'_{i,1}) \ \forall i. \tag{12}$$

To prove Equation 12, notice that

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

Because the players' valuations are always 0 in $\mathcal{T}'$, by another induction we have that, for any i, k', ℓ,

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

Thus

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

Accordingly, we have $a_i^k \in RAT_i^k(t'_{i,0}) \subseteq RAT_i^{k-1}(t'_{i,0}) = RAT_i^{k-1}(t'_{i,1})$ for each i, and Equation 12 holds. Therefore Statement 11 also holds, and so does Statement 10. Combining Equation 9 and Statement 10, we have that Statement 8 holds, a contradiction. Thus Theorem 2 holds for $n = 2$ and $k \geq 1$.

The analysis is very similar for the degenerated case where $n = 2$ and $k = 0$. Indeed, we consider the context $C = (2, V, \mathcal{T}, \tau)$ with $\mathcal{T} = (T, \Theta, \nu, B)$ defined as follows: for each player i,

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

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

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

Because $\hat{M}$ is IIR, in auction $(C', \hat{M})$ there exists an action profile a such that $u_i(a, 0) \geq 0$ for each i. Thus $rev(\hat{M}(a)) \leq 0 < V - c = G^0(C) - c$. Because $a \in A = RAT^0(\tau)$, $\hat{M}$ cannot level-0 rationally[2] implement $G^0 - c$. In sum, Theorem 2 holds for $n = 2$.

Finally, for $n > 2$, we construct the desired type structures (and contexts) essentially by adding dummy players to the type structures $\mathcal{T}$ and $\mathcal{T}'$. More precisely, the n-player type structure $\hat{\mathcal{T}} = (\hat{T}, \Theta, \hat{\nu}, \hat{B})$ is defined as follows:

- $\forall i \in \{1, 2\}$, $\hat{T}_i = T_i$;
- $\forall i \notin \{1, 2\}$, $\hat{T}_i = \{\hat{t}_i\}$;
- $\forall i \in \{1, 2\}$, $\hat{\nu}_i(t_i) = \nu_i(t_i)$ for any $t_i \in \hat{T}_i$;
- $\forall i \notin \{1, 2\}$, $\hat{\nu}_i(\hat{t}_i) = 0$;
- $\forall i \in \{1, 2\}$, $\hat{B}_i(t_i) = B_i(t_i) \times \{\hat{t}_{-\{1,2\}}\}$ for all $t_i \in \hat{T}_i$;
- $\forall i \notin \{1, 2\}$, $\hat{B}_i(\hat{t}_i) = \{(t_{1,0}, t_{2,0}, \hat{t}_{-\{1,2,i\}})\}$.

In the context $\hat{C} = (n, V, \hat{\mathcal{T}}, \hat{\tau})$, we let $\hat{\tau} = (\tau, \hat{t}_{-\{1,2\}})$. The auxiliary type structure $\hat{\mathcal{T}}' = (\hat{T}', \Theta, \hat{\nu}', \hat{B}')$ is constructed from $\mathcal{T}'$ in the same way, and so is the auxiliary context $\hat{C}'$. The analysis is essentially the same, and thus omitted.

In sum, Theorem 2 holds. ■

S3 Variants of Mechanism M

Discrete versus Continuous Valuation Space From the examples that we have discussed in the main paper and the analysis in the Supplementary Material, it is not hard to see that the revenue guarantee of our mechanism is facilitated by the fact that the valuation space and thus the action space of the mechanism are discrete—so that a player i who wants to increase v_i must increase it by at least 1 and the bigger reward he gets from this offsets the smaller reward due to the possible increase of ℓ_i. If the values can be reals and the mechanism allows the v_i's to be reals, then in the second round of elimination in the first example of Section 5.2, player 1 wants to announce

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

v_1 smaller than but arbitrarily close to 200, believing that player 2 will announce $v_2 \geq 200$ and $\ell_2 = 0$. However, any action (ℓ_1, v_1) of player 1 with $v_1 < 200$ is dominated by $(0, v_1 + \frac{200-v_1}{2})$ and thus should be eliminated. The limit, $(0, 200)$, is not dominated, but it does not dominate the actions (ℓ_1, v_1) with $v_1 < 200$ either.

More generally, with a continuous valuation space Theorem 1 remains true under a slightly different analysis, but our mechanism becomes *unbounded* [13]: the dominated strategies are not dominated by any of the surviving ones. Following [13], we focus on bounded mechanisms, and that is why we only consider discrete valuation spaces in our model. It is an interesting open problem whether there exists a bounded mechanism for continuous valuation spaces that leads to Theorem 1.

Finite versus Infinite Action Spaces We would like to point out that the finite valuation bound V and level bound K are needed only to ensure that our mechanism has a finite action space. We impose this restriction because our epistemic characterization in Section S4 of level-k rationality (i.e., by means of an iterated deletion procedure) only applies to finite games, similar to many other characterizations of higher-level rationality [8, 6, 5, 4, 16, 11].

We note, however, that the analysis in Theorem 1 (which focuses only on the set of actions surviving the iterated deletion procedure) applies also to a variant of our mechanism M without these finite bounds: namely, a mechanism M' defined identically to M except that each player i announces $(i, \ell_i, v_i) \in \{i\} \times \mathbb{Z}^+ \times \mathbb{Z}^+$, where $\mathbb{Z}^+$ is the set of nonnegative integers. We emphasize that this holds as long as we consider a *finite* rationality level k, as we next discuss.

Infinitely Rational Players With a finite action space, our mechanism M can only elicit the players' beliefs up to level K, even when they are infinitely rational. As mentioned above, the variant M' can elicit the players' beliefs up to any finite level k, as long as the players are level-$(k+1)$ rational. When the players are infinitely rational—that is, level-k rational for every $k \geq 0$, consider $g_i^\infty = \max_k g_i^k$ for each i and let G^∞ be the second highest of the g_i^∞'s.

As long as either the type space or the valuation space is finite, each g_i^∞ is finite and can be attained at some finite belief level k_i. Roughly speaking, g_i^∞ is the highest "rumored" valuation according to player i's beliefs and k_i is the "closeness" of the rumor. In this case, the variant M' leverages the players' infinitely high rationality levels *without having any information about the g_i^∞'s or the k_i's.* Allegedly, each player i announces (a) $v_i = g_i^\infty$, the highest value v such that i believes *"there exists some player who believes"* ... some player values the good v, and (b) $\ell_i = k_i$, the smallest level of beliefs about beliefs needed to attain v_i. The analysis is almost the same. In particular, M' guarantees the revenue benchmark $G^\infty - \varepsilon$ under common belief of rationality.

If both the valuation space and the type space are infinite, then there exist contexts where, for each player i, g_i^k goes to infinity as k goes to infinity. In this case, there is no action profile consistent with common belief of rationality in M', since each action (i, ℓ_i, v_i) will be eliminated in some round k_i where $g_i^{k_i}$ exceeds v_i.

Our Mechanism Under Different Solution Concepts Although our solution concept only requires a very weak notion of rationality, it is interesting to consider how the mechanism behaves under other solution concepts that impose stronger assumptions about the players' rationality and/or their beliefs about each other's types. For example, following [3], sufficient conditions (which are tight in some sense) for Nash equilibrium require that the true type profile is mutual knowledge among the players, which implies the players have correct beliefs under our model. When the players do have correct beliefs, it is easy to see that our elimination procedure preserves all

(including mixed) equilibrium actions, since it only eliminates actions that are strictly dominated. Thus the set of Nash equilibria actually implements the benchmark G^∞ as defined above. A characterization for the structure of Nash equilibria in our mechanism remains unknown (e.g., for many type structures there is no pure Nash equilibrium, since the winner can improve his utility by bidding a higher value to get a higher reward). Such a characterization, although interesting to explore, is beyond the scope of this paper.

When additional probabilistic structure is added to the type structure, one can consider a stronger notion of rationality based on the players' *expected* utilities, and define corresponding iterated elimination of dominated actions (see, e.g., [1]). However, a probabilistic structure must be *consistent* with the players' possibilistic beliefs: namely, a player never assigns positive probability to a type that he believes to be impossible according to his possibilistic beliefs. It is easy to see that for any consistent probabilistic structure, any action that is eliminated under our solution concept must also be eliminated based on the stronger notion of rationality. Thus our mechanism continues to implement our benchmarks and Theorem 1 continues to hold. Moreover, when there is a common prior over the type structure, our mechanism implements the benchmark G^∞ under Bayesian Nash equilibria, although the structure of such equilibria has not been characterized yet.

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

S4 Characterization of Level-k Rationality

We consider rationality and rationalizability for finite normal-form games of incomplete information in which the players have *possibilistic* beliefs about their opponents. In this setting, we prove that the actions consistent with the players being level-k rational coincide with the actions surviving a natural k-step iterated elimination procedure. We view the latter actions as the (level-k) rationalizable ones in our possibilistic setting. Section S4.2 and Definitions S4 and S5 are the main conceptual novelty in this Supplement (even though some notions in Section S4 are similar to those in [10], the characterization of level-k rationality and the connection between possibilistic structures and type structures are quite nontrivial).

Rationalizability was defined by Pearce [14] and Bernheim [7] for complete-information settings. Our iterated elimination procedure is similar to that proposed by Dekel, Fudenberg, and Morris [8] and by Bergemann and Morris [6] in a Bayesian setting. For other iterated elimination procedures and corresponding notions of rationalizability in Bayesian settings, see Brandenburger and Dekel [5], Tan and Werlang [15], Battigalli and Siniscalchi [4], Ely and Peski [9], Weinstein and Yildiz [16], and Halpern and Pass [11].

S4.1 Possibilistic Structures and Rationality Models

Given an n-player normal-form game Γ, let A_i be the finite set of pure actions of player i in Γ and $A = A_1 \times \cdots \times A_n$. To model the players' uncertainty about each other's utility and action in Γ, we consider a possibilistic version of Harsanyi's type structure [12].

Definition S1. A *possibilistic structure* $\mathcal{G}$ for Γ is a tuple of profiles, $\mathcal{G} = (T, u, B, \boldsymbol{s})$, where for each player i,

- T_i is a finite set, the set of i's possible *types*;
- $u_i : A \times T \to \mathbb{R}$ is i's *utility function*;
- $B_i : T_i \to 2^{T_{-i}}$ is i's *belief correspondence*; and
- $\boldsymbol{s}_i : T_i \to A_i$ is i's *strategy function*.

A possibilistic structure 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. Moreover, each utility function u_i has domain $A \times T$ rather than $A \times T_i$. This enables us to deal with interdependent-type settings as well.

Below, we define the players' rationality, higher-level rationality, and common belief of rationality, in the same way as Aumann [2]. The notions we use and the basic properties we prove about them can be considered as the possibilistic analog of those in [10].

Definition S2. Let $\mathcal{G} = (T, u, B, \boldsymbol{s})$ be a possibilistic structure for Γ and t be a type profile in T. Player i is *rational* at t_i if, for every action a_i' of i, there exists $t'_{-i} \in B_i(t_i)$ such that

$$u_i((\boldsymbol{s}_i(t_i), \boldsymbol{s}_{-i}(t'_{-i})), (t_i, t'_{-i})) \geq u_i((a_i', \boldsymbol{s}_{-i}(t'_{-i})), (t_i, t'_{-i})).$$

Player i is *rational at* t if he is rational at t_i.

Based on this definition, we define the following events.

- Let $RAT_i = \{t \in T |\ i \text{ is rational at } t\}$ be the event that player i is rational.
- For any event $E \subseteq T$, let $\mathbf{B}_i(E) = \{t \in T |\ (t_i, t'_{-i}) \in E\ \forall t'_{-i} \in B_i(t_i)\}$ be the event that player i believes that E occurs.
- Let $RAT_i^0 = T$ be the event that player i is level-0 rational (namely, irrational), and for any $k \geq 1$, let $RAT_i^k = RAT_i \cap \mathbf{B}_i(\cap_{j \neq i} RAT_j^{k-1})$ be the event that player i is level-k rational. Clearly, $RAT_i^1 = RAT_i \cap \mathbf{B}_i(\cap_{j \neq i} RAT_j^0) = RAT_i \cap \mathbf{B}_i(T) = RAT_i \cap T = RAT_i$. That is, being level-1 rational is equivalent to being rational.
- For any $k \geq 0$, let $RAT^k = \cap_i RAT_i^k$ be the event that every player is level-k rational, and let $RAT = RAT^1$ be the event that every player is rational.
- For any event $E \subseteq T$, let $\mathbf{EB}^0(E) = E$, $\mathbf{EB}^1(E) = \mathbf{EB}(E) = \cap_i \mathbf{B}_i(E)$ be the event that every player believes that E occurs, and $\mathbf{EB}^k(E) = \mathbf{EB}(\mathbf{EB}^{k-1}(E))$ for any $k \geq 2$.
- Let $\mathbf{CB}(RAT) = \cap_{k \geq 0} \mathbf{EB}^k(RAT)$ be the event that the players have common belief of rationality.

Definition S3. For any $t \in T$ and $k \geq 0$, player i is *level-k rational at* t if $t \in RAT_i^k$. For any $t_i \in T_i$, player i is *level-k rational at* t_i if there exists $t_{-i} \in T_{-i}$ such that i is level-k rational at (t_i, t_{-i}). For any $t \in T$, the players have *common belief of rationality at* t if $t \in \mathbf{CB}(RAT)$.

Notice that whether player i is level-k rational or not at t solely depends on t_i and player i's belief hierarchy at t_i, and does not depend on t_{-i} at all. Thus it is immediately clear that

($*$) *Player i is level-k rational at t_i if and only if, for all $t_{-i} \in T_{-i}$, player i is level-k rational at (t_i, t_{-i}).*

Basic Properties of Our Model The following three properties (which are standard in epistemic game theory) help understanding our model.

Property S1. *For any player i and any $k \geq 1$, $RAT_i^k \subseteq RAT_i^{k-1}$.*

Property S2. *For any player i and any $k \geq 1$, $RAT_i^k = RAT_i \cap \mathbf{B}_i(\cap_j RAT_j^{k-1})$.*

Property S3. $\mathbf{CB}(RAT) = \cap_{k \geq 0} \cap_{i \in [n]} RAT_i^k$.

In particular, Property S1 shows that the players' higher levels of rationality are nested. Property S2 is a trivial corollary of Property S1 and is also a natural way to think about level-k rationality—that is, being level-k rational is equivalent to being rational and believing that *every player* is level-$(k-1)$ rational. It will be used in the proof of Theorem S2. Finally, Property S3 provides an alternative definition for common belief of rationality. Its proof relies on Properties S1 and S2, and it will also be used in the proof of Theorem S2. To prove these properties, we first state without proofs the following simple observations.

1. For any player i, $RAT_i = \mathbf{B}_i(RAT_i)$.

   That is, a rational player believes that he is rational.

2. For any player i and any $k \geq 0$, $RAT_i^k = \mathbf{B}_i(RAT_i^k)$.

   That is, a level-k rational player believes that he is level-k rational.

Proof of Property S1. By induction on k. For $k = 1$, $RAT_i^1 \subseteq T = RAT_i^0$. For $k > 1$, by the induction hypothesis we have $RAT_j^{k-1} \subseteq RAT_j^{k-2}$ for each j, thus $\mathbf{B}_i(\cap_{j \neq i} RAT_j^{k-1}) \subseteq \mathbf{B}_i(\cap_{j \neq i} RAT_j^{k-2})$. Accordingly, $RAT_i^k = RAT_i \cap \mathbf{B}_i(\cap_{j \neq i} RAT_j^{k-1}) \subseteq RAT_i \cap \mathbf{B}_i(\cap_{j \neq i} RAT_j^{k-2}) = RAT_i^{k-1}$, as desired. □

Proof of Property S2. Since $RAT_i^k = RAT_i \cap \mathbf{B}_i(\cap_{j \neq i} RAT_j^{k-1})$ by definition, $RAT_i^k \subseteq RAT_i^{k-1}$ by Property S1, and $RAT_i^{k-1} = \mathbf{B}_i(RAT_i^{k-1})$ by Observation 2, we have

$$\begin{aligned} RAT_i^k &= RAT_i \cap \mathbf{B}_i(\cap_{j \neq i} RAT_j^{k-1}) \cap RAT_i^{k-1} = RAT_i \cap \mathbf{B}_i(\cap_{j \neq i} RAT_j^{k-1}) \cap \mathbf{B}_i(RAT_i^{k-1}) \\ &= RAT_i \cap \mathbf{B}_i(\cap_j RAT_j^{k-1}), \end{aligned}$$

as desired. □

Proof of Property S3. We show by induction that, for any $k \geq 1$, $\cap_i RAT_i^k = \mathbf{EB}^{k-1}(RAT)$. For $k = 1$, $\cap_i RAT_i^1 = RAT^1 = RAT = \mathbf{EB}^0(RAT)$ as desired. For $k > 1$,

$$\begin{aligned} \cap_i RAT_i^k &= \cap_i \left(RAT_i \cap \mathbf{B}_i(\cap_j RAT_j^{k-1}) \right) = \cap_i \left(\mathbf{B}_i(RAT_i) \cap \mathbf{B}_i(\cap_j RAT_j^{k-1}) \right) \\ &= \cap_i \left(\mathbf{B}_i((RAT_i^1 \cap RAT_i^{k-1}) \cap (\cap_{j \neq i} RAT_j^{k-1})) \right) \\ &= \cap_i \mathbf{B}_i(RAT_i^{k-1} \cap (\cap_{j \neq i} RAT_j^{k-1})) = \cap_i \mathbf{B}_i(\cap_j RAT_j^{k-1}) \\ &= \mathbf{EB}(\cap_j RAT_j^{k-1}) = \mathbf{EB}(\mathbf{EB}^{k-2}(RAT)) = \mathbf{EB}^{k-1}(RAT). \end{aligned}$$

The first equality is due to Property S2, the second to Observation 1, the fourth to Property S1, and the seventh to the induction hypothesis. Since $\cap_i RAT_i^0 = T$, we have

$$\cap_{k \geq 0} \cap_i RAT_i^k = \cap_{k \geq 1} \cap_i RAT_i^k = \cap_{k \geq 0} \mathbf{EB}^k(RAT) = \mathbf{CB}(RAT),$$

as desired. □

S4.2 Type Structures and Iterated Elimination of Strictly Dominated Actions

In many scenarios, the players' beliefs about each other's (payoff) types are given exogenously, and they reason about each other's actions based on their beliefs about types. To model this kind of information structure and reasoning procedure, we define *type structures*: a type structure $\mathcal{T}$ for Γ is a tuple of profiles, $\mathcal{T} = (T, u, B)$, where T, u, B are as defined in a possibilistic structure for Γ. Thus a type structure can be considered as a possibilistic structure with the strategy function removed.

Definition S4. A possibilistic structure $\mathcal{G} = (T, u, B, \boldsymbol{s})$ for Γ is *consistent* with a type structure $\mathcal{T}' = (T', u', B')$ for Γ if there exists a profile of functions ψ with $\psi_i : T_i \to T_i'$ $\forall i$ such that,

- $\forall i$ and $\forall t \in T$, $u_i(\cdot; t) = u_i'(\cdot; \psi(t))$; and
- $\forall i$ and $\forall t_i \in T_i$, $\psi_{-i}(B_i(t_i)) = B_i'(\psi_i(t_i))$.

We refer to such a ψ as a *consistency mapping*.

The notion of consistency captures that introducing actions into the picture does not cause the players to change their beliefs about *types*, but causes them to form additional beliefs about *actions*.

Illustratively, both possibilistic structures and type structures can be represented by directed graphs, with nodes corresponding to the players' types and edges corresponding to their beliefs. The only difference is that, in a possibilistic structure, each node is also associated with an action.

Example Consider a revised version of the BoS game, where player 1 has a unique type t_1 and player 2 has two types t_2 and t_2' —whether he wants to meet or avoid player 1. The players' utilities are specified in Figure 3.

	B	S
B	2,1	0,0
S	0,0	1,2

(a) Utilities under (t_1, t_2)

	B	S
B	2,0	0,2
S	0,1	1,0

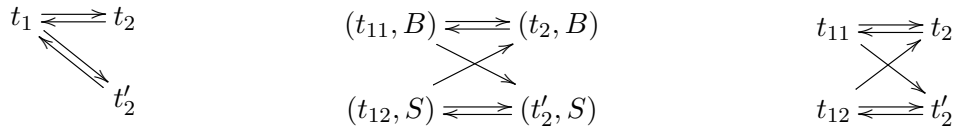
(b) Utilities under (t_1, t_2')

Figure 3: A revised BoS game.

Figure 4a provides an elementary type structure $\mathcal{T}'$ for the revised BoS game, where player 1 believes that player 2's type can be either t_2 or t_2' and player 2 believes that player 1's (unique) type is t_1. Figure 4b provides an elementary possibilistic structure $\mathcal{G}$ consistent with $\mathcal{T}'$. Here player 1's two types t_{11} and t_{12} induce the same utility function but different actions for him, and under both types player 1 believes that player 2 will use action B under type t_2 and S under t_2'. The type structure $\mathcal{T}$ obtained from $\mathcal{G}$ by removing the actions is then illustrated in Figure 4c. It is immediate to see that the consistency mapping $\psi = (\psi_1, \psi_2)$ is such that ψ_1 maps both t_{11} and t_{12} to t_1, and ψ_2 maps t_2 to t_2 and t_2' to t_2'. Indeed, under such mapping, the utilities are preserved and "the belief correspondence and ψ commute."

We now define rationality for type structures.

Definition S5. Given a type structure $\mathcal{T} = (T, u, B)$ for Γ, for any player i, type $t_i \in T_i$, action a_i, and integer $k \geq 0$, a_i is *consistent with level-k rationality for* t_i if there exists a possibilistic structure $\mathcal{G} = (T', u', B', \boldsymbol{s})$ and a type $t_i' \in T_i'$, such that $\mathcal{G}$ is consistent with $\mathcal{T}$ under a consistency mapping ψ, $\psi_i(t_i') = t_i$, $\boldsymbol{s}_i(t_i') = a_i$, and i is level-k rational at t_i'.



(a) Type structure $\mathcal{T}'$ (b) Possibilistic structure $\mathcal{G}$. (c) Type structure $\mathcal{T}$

Figure 4: A type structure and a consistent possibilistic structure

Action a_i is *consistent with common belief of rationality for* t_i if there exists a possibilistic structure $\mathcal{G} = (T', u', B', \boldsymbol{s})$ and a type profile $t' \in T'$, such that $\mathcal{G}$ is consistent with $\mathcal{T}$ under a consistency mapping ψ, $\psi_i(t'_i) = t_i$, $\boldsymbol{s}_i(t'_i) = a_i$ and the players have common belief of rationality at t'.

Slightly abusing notations, we denote by $RAT_i^k(t_i)$ the set of actions consistent with level-k rationality for t_i and by $RAT_i(t_i)$ the set of actions consistent with common belief of rationality for t_i. Notice that our concept of consistency with level-k rationality or common belief of rationality is called *rationalizability* in other studies; see [4]. Next we define an iterated elimination procedure for refining the players' actions, and use it to characterize actions that are consistent with level-k rationality or common belief of rationality.

Definition S6. Let $\mathcal{T} = (T, u, B)$ be a type structure for Γ. For each player i, type $t_i \in T_i$, and integer $k \geq 0$, we define $NSD_i^k(t_i)$, *the set of actions surviving* k*-round elimination of strictly dominated actions for* t_i, inductively as follows:

- $NSD_i^0(t_i) = A_i$.
- For each $k \geq 1$ and each $a_i \in NSD_i^{k-1}(t_i)$, $a_i \in NSD_i^k(t_i)$ if there does not exist an alternative action $a'_i \in NSD_i^{k-1}(t_i)$ such that $\forall t_{-i} \in B_i(t_i)$ and $\forall a_{-i} \in NSD_{-i}^{k-1}(t_{-i})$,

$$u_i((a'_i, a_{-i}), (t_i, t_{-i})) > u_i((a_i, a_{-i}), (t_i, t_{-i})),$$

  where $NSD_{-i}^{k-1}(t_{-i}) = \times_{j \neq i} NSD_j^{k-1}(t_j)$.

In the definition for $NSD_i^k(t_i)$, if the required action a'_i does exist, we say that a_i is strictly dominated (by a'_i) for t_i over level-$(k-1)$ surviving actions. It is easy to see that defining $NSD_i^k(t_i)$ by eliminating strictly dominated actions from $NSD_i^{k-1}(t_i)$ is the same as defining it by eliminating strictly dominated actions from A_i. Indeed, we have the following lemma, whose proof has been omitted.

Lemma S1. *For any* $k \geq 1$ *and* $a_i \in A_i$, $a_i \in NSD_i^k(t_i)$ *if and only if there does not exist an alternative action* $a'_i \in A_i$ *such that* a_i *is strictly dominated by* a'_i *for* t_i *over level-*$(k-1)$ *surviving actions.*

Given player i's knowledge about $\mathcal{T}$, he can iteratively compute $NSD_i^k(t_i)$ for any t_i and k. Since both the game Γ and the type structure $\mathcal{T}$ are finite, the elimination procedure ends for all types of all players after some round K when no action is strictly dominated over level-$(K-1)$ surviving actions. Letting

$$NSD_i(t_i) = \cap_{k \geq 0} NSD_i^k(t_i),$$

we have $NSD_i(t_i) = NSD_i^K(t_i) \neq \emptyset$. We say that an action a_i survives iterated elimination of strictly dominated actions for t_i if $a_i \in NSD_i(t_i)$. Following [4], we refer to $NSD_i^k(t_i)$ as the set of *level-k rationalizable* actions for t_i, and to $NSD_i(t_i)$ as the set of *rationalizable* actions for t_i.

An immediate consequence of Lemma S1 is the following lemma, stated without proof.

Lemma S2. *For any $k \geq 1$ and $a_i \in A_i$, $a_i \in NSD_i^k(t_i)$ if and only if there exists $B_i' \subseteq B_i(t_i)$ and $Z_{-i}(t_{-i}) \subseteq NSD_{-i}^{k-1}(t_{-i})$ for each $t_{-i} \in B_i'$, such that, for each $a_i' \in A_i$, there exists $t_{-i} \in B_i'$ and $a_{-i} \in Z_{-i}(t_{-i})$ with*

$$u_i((a_i, a_{-i}), (t_i, t_{-i})) \geq u_i((a_i', a_{-i}), (t_i, t_{-i})).$$

Intuitively, a_i survives k-round elimination if, given i's belief that other players' types are among (some subset of) $B_i(t_i)$ and they use (some subset of) actions that survive $(k-1)$-round elimination, no other action *according to i's belief* can lead to higher utility than what he gets by using a_i. Lemma S2 is a possibilistic analog of Pearce's lemma [14] which, in probabilistic models, relates best responses and rationalizability to strict dominance. Note that whereas in the possibilistic case (which is what we consider) the proof is trivial, Pearce's original lemma for the probabilistic case requires additional work.

S4.3 Characterizing Level-k Rationality and Common Belief of Rationality

Theorem S1. *Given a type structure $\mathcal{T} = (T, u, B)$ for Γ, for any player i, type t_i, action a_i, and integer $k \geq 0$, a_i is consistent with level-k rationality for t_i if and only if $a_i \in NSD_i^k(t_i)$; that is, $RAT_i^k(t_i) = NSD_i^k(t_i)$.*

Proof. We first prove the "only if" direction. Assuming a_i is consistent with level-k rationality for t_i, we prove $a_i \in NSD_i^k(t_i)$ by induction on k. For $k = 0$, the property trivially holds since $NSD_i^0(t_i) = A_i$ by definition.

For $k > 0$, by Definition S5 there exists a possibilistic structure $\mathcal{G} = (T', u', B', \boldsymbol{s})$ and a type $t_i' \in T_i'$, such that $\mathcal{G}$ is consistent with $\mathcal{T}$ under a consistency mapping ψ, $\psi_i(t_i') = t_i$, $\boldsymbol{s}_i(t_i') = a_i$, and i is level-k rational at t_i'.

By Definition S3 and Property $(*)$, player i being level-k rational at t_i' implies: (a) i is rational at t_i'; and (b) for each type subprofile $t_{-i}' \in B_i'(t_i')$, we have $(t_i', t_{-i}') \in \cap_{j \neq i} RAT_j^{k-1}$. According to (a) and Definition S2, for each action $a_i' \in A_i$ there exists $t_{-i}' \in B_i'(t_i')$ such that

$$u_i'((a_i, \boldsymbol{s}_{-i}(t_{-i}')), (t_i', t_{-i}')) \geq u_i'((a_i', \boldsymbol{s}_{-i}(t_{-i}')), (t_i', t_{-i}')). \tag{13}$$

According to (b), for each $t_{-i}' \in B_i'(t_i')$ and each $j \neq i$, player j is level-$(k-1)$ rational at t_j'. By Definition S5, $\boldsymbol{s}_j(t_j')$ is consistent with level-$(k-1)$ rationality for $\psi_j(t_j')$ and thus, by the induction hypothesis,

$$\boldsymbol{s}_j(t_j') \in NSD_j^{k-1}(\psi_j(t_j')). \tag{14}$$

For each $t_{-i} \in B_i(t_i)$, let $Z_{-i}(t_{-i}) = \boldsymbol{s}_{-i}(\psi_{-i}^{-1}(t_{-i}))$. Because $\psi_{-i}(B_i'(t_i')) = B_i(t_i)$, $Z_{-i}(t_{-i}) \neq \emptyset$. By Equation 14,

$$Z_{-i}(t_{-i}) \subseteq NSD_{-i}^{k-1}(t_{-i}).$$

For each $a_i' \in A_i$, let $t_{-i}' \in B_i'(t_i')$ be such that Equation 13 holds, $t_{-i} = \psi_{-i}(t_{-i}')$, and $a_{-i} = \boldsymbol{s}_{-i}(t_{-i}')$. Accordingly, $a_{-i} \in Z_{-i}(t_{-i})$. Since $u_i(\cdot; (t_i, t_{-i})) = u_i'(\cdot; (t_i', t_{-i}'))$, Equation 13 implies

$$u_i((a_i, a_{-i}), (t_i, t_{-i})) \geq u_i((a_i', a_{-i}), (t_i, t_{-i})).$$

By Lemma S2, we have $a_i \in NSD_i^k(t_i)$, concluding the proof of the "only if" direction.

Now we prove the "if" direction. By definition, proving this direction is equivalent to proving that, if $a_i \in NSD_i^k(t_i)$, then there exists a possibilistic structure $\mathcal{G} = (T', u', B', \boldsymbol{s})$ for Γ and a type $t_i' \in T_i'$ such that $\mathcal{G}$ is consistent with $\mathcal{T}$ under a consistency mapping ψ, $\psi_i(t_i') = t_i$, $\boldsymbol{s}_i(t_i') = a_i$, and i is level-k rational at t_i'. Notice that $\mathcal{G}$, t_i', and ψ may depend on k, i, t_i, and a_i.

In fact, we shall prove a stronger statement. Namely, for each k, there exists a *universal* possibilistic structure $\mathcal{G} = (T', u', B', \boldsymbol{s})$ for Γ, consistent with $\mathcal{T}$ under a consistency mapping ψ, such that, for *every* player i, type $t_i \in T_i$, action a_i, and nonnegative integer $k' \leq k$,

if $a_i \in NSD_i^{k'}(t_i)$ then there exists a type $t_i' \in T_i'$ such that

$$\psi_i(t_i') = t_i, \quad \boldsymbol{s}_i(t_i') = a_i \quad \text{and} \quad i \text{ is level-}k' \text{ rational at } t_i', \tag{15}$$

which implies that a_i is consistent with level-k' rationality for t_i'.

We define $\mathcal{G}$ as follows: for each player i,

- $T_i' = \Big\{(t_i, k', a_i) : t_i \in T_i, k' \in \{0, \ldots, k\}, a_i \in NSD_i^{k'}(t_i)\Big\}$;
- for each type profile $t' \in T'$, letting $t \in T$ be the type profile obtained by projecting each t_j' to its first component, $u_i'(\cdot; t') = u_i(\cdot; t)$;
- for each type $t_i' = (t_i, k', a_i)$, $\boldsymbol{s}_i(t_i') = a_i$; and
- for each type $t_i' = (t_i, k', a_i)$ and type subprofile $t_{-i}' \in T_{-i}'$, $t_{-i}' \in B_i'(t_i')$ if and only if there exist $t_{-i} \in B_i(t_i)$ and $a_{-i} \in NSD_{-i}^{\max\{k'-1,0\}}(t_{-i})$ such that $t_j' = (t_j, \max\{k'-1, 0\}, a_j)$ for all $j \neq i$.

It is easy to check that $\mathcal{G}$ is consistent with $\mathcal{T}$ under the consistency mapping ψ where $\psi_i(t_i, k', a_i) = t_i$ for each player i and type $(t_i, k', a_i) \in T_i'$.

We now prove by induction on k' that, for any $i, t_i \in T_i$, and $a_i \in NSD_i^{k'}(t_i)$, player i is level-k' rational at $t_i' = (t_i, k', a_i)$. For $k' = 0$, since $RAT_i^0 = T$ by definition, it trivially holds that player i is level-0 rational at t_i'.

For $k' > 0$, for any $t_{-i}' = (t_j, k'-1, a_j)_{j \neq i} \in B_i'(t_i')$, by construction we have $t_{-i} \in B_i(t_i)$ and $a_{-i} \in NSD_{-i}^{k'-1}(t_{-i})$. By the hypothesis induction, for any player $j \neq i$, j is level-$(k'-1)$ rational at t_j' and thus at (t_i', t_{-i}'). Therefore

$$(t_i', t_{-i}') \in \cap_{j \neq i} RAT_j^{k'-1}.$$

Since this is true for any $t_{-i}' \in B_i'(t_i')$, we have

$$(t_i', t_{-i}') \in \mathbf{B}_i(\cap_{j \neq i} RAT_j^{k'-1})$$

for any $t_{-i}' \in B_i'(t_i')$, as again whether player i believes some event or not only depends on t_i' and not t_{-i}'.

Since $a_i \in NSD_i^{k'}(t_i)$, by Lemma S1 we have that, for any $a_i' \in A_i$, there exist $t_{-i} \in B_i(t_i)$ and $a_{-i} \in NSD_{-i}^{k'-1}(t_{-i})$ such that

$$u_i((a_i, a_{-i}), (t_i, t_{-i})) \geq u_i((a_i', a_{-i}), (t_i, t_{-i})).$$

Letting $t_{-i}' = (t_j, k'-1, a_j)_{j \neq i}$, we have $t_{-i}' \in B_i'(t_i')$, $\psi(t_i', t_{-i}') = (t_i, t_{-i})$, $\boldsymbol{s}_i(t_i') = a_i$, and $\boldsymbol{s}_{-i}(t_{-i}') = a_{-i}$. Thus

$$u_i'((\boldsymbol{s}_i(t_i'), \boldsymbol{s}_{-i}(t_{-i}')), (t_i', t_{-i}')) \geq u_i'((a_i', \boldsymbol{s}_{-i}(t_{-i}')), (t_i', t_{-i}')).$$

Accordingly, player i is rational at t'_i and $(t'_i, t'_{-i}) \in RAT_i$ for any $t'_{-i} \in B'_i(t'_i)$. By definition, $(t'_i, t'_{-i}) \in RAT_i \cap \mathbf{B}_i(\cap_{j \neq i} RAT_j^{k'-1})$ for any $t'_{-i} \in B'_i(t'_i)$, and thus i is level-k' rational at t'_i. This concludes the induction step and the proof of Statement (15). Therefore the "if" direction holds, concluding the proof of Theorem S1. ■

Similarly, we characterize common belief of rationality in our model by the following theorem.

Theorem S2. *Given a type structure $\mathcal{T} = (T, u, B)$ for Γ, for any player i, type t_i and action a_i, a_i is consistent with common belief of rationality for t_i if and only if $a_i \in NSD_i(t_i)$: that is, $RAT_i(t_i) = NSD_i(t_i)$.*

Proof. We first prove the "only if" direction. Assume a_i is consistent with common belief of rationality for t_i. By Definition S5, there exist a possibilistic structure $\mathcal{G} = (T', u', B', \boldsymbol{s})$ and a type profile $t' \in T'$, such that $\mathcal{G}$ is consistent with $\mathcal{T}$ under a consistency mapping ψ, $\psi_i(t'_i) = t_i$, $\boldsymbol{s}_i(t'_i) = a_i$, and $t' \in \mathbf{CB}(RAT)$.

By Property S3, for any $k \geq 0$, $t' \in RAT_i^k$ and player i is level-k rational at t'_i. Thus, by Definition S5, a_i is consistent with level-k rationality for t_i. By Theorem S1, $a_i \in NSD_i^k(t_i)$ for any $k \geq 0$. Thus $a_i \in NSD_i(t_i)$ and the "only if" direction holds.

The "if" direction holds from the following lemma, which we prove separately.

Lemma S3. *There exists a* universal *possibilistic structure $\mathcal{G} = (T', u', B', \boldsymbol{s})$ for Γ, consistent with $\mathcal{T}$ under a consistency mapping ψ, such that*

(1) $\mathbf{CB}(RAT) = T'$ *—that is, common belief of rationality holds everywhere, and*

(2) for every *player i, type $t_i \in T_i$, and action $a_i \in NSD_i(t_i)$, there exists a type $t'_i \in T'_i$ such that $\psi_i(t'_i) = t_i$ and $\boldsymbol{s}_i(t'_i) = a_i$.*

Indeed, Lemma S3 implies that, for any i, t_i, and $a_i \in NSD_i(t_i)$, a_i is consistent with common belief of rationality for t_i.

In sum, Theorem S2 holds. ■

Proof of Lemma S3. Similarly to the second part of the proof of Theorem S1, we construct structure $\mathcal{G}$ as follows: for each player i,

- $T'_i = \{(t_i, a_i) : t_i \in T_i, a_i \in NSD_i(t_i)\}$;
- for each type profile $t' \in T'$, letting $t \in T$ be the type profile obtained by projecting each t'_j to its first component, $u'_i(\cdot; t') = u_i(\cdot; t)$;
- for each $t'_i = (t_i, a_i) \in T'_i$, $\boldsymbol{s}_i(t'_i) = a_i$; and
- for each $t'_i = (t_i, a_i) \in T'_i$ and $t'_{-i} \in T'_{-i}$, $t'_{-i} \in B'_i(t'_i)$ if and only if there exist $t_{-i} \in B_i(t_i)$ and $a_{-i} \in NSD_{-i}(t_{-i})$ such that $t'_j = (t_j, a_j)$ for all $j \neq i$.

Below, we only show part (1) of Lemma S3, as part (2) holds by construction.

By Property S3, to show $\mathbf{CB}(RAT) = T'$ it suffices to show $\cap_{j \in [n]} RAT_j^k = T'$ for every $k \geq 0$. We proceed by induction. For $k = 0$, by definition, $\cap_{j \in [n]} RAT_j^0 = \cap_{j \in [n]} T' = T'$.

For $k > 0$, it suffices to show $RAT_i^k = T'$ for each player i. Arbitrarily fixing a player i, by the induction hypothesis we have $\cap_{j \in [n]} RAT_j^{k-1} = T'$, and thus $\mathbf{B}_i(\cap_{j \in [n]} RAT_j^{k-1}) = T'$ as well. By Property S2, it is left to show $RAT_i = T'$, or equivalently, player i is rational at every type $t'_i \in T'_i$.

Arbitrarily fix a type $t_i' = (t_i, a_i)$ and an action a_i' of i. Since $a_i \in NSD_i(t_i)$, by definition and by Lemma S1 we have that a_i is not strictly dominated by a_i' for t_i over level-ℓ surviving actions for any $\ell \geq 0$. In particular, a_i is not strictly dominated by a_i' for t_i over level-K surviving actions, where K is such that the elimination procedure ends after round K for all types of all players. That is, there exists $t_{-i} \in B_i(t_i)$ and $a_{-i} \in NSD_{-i}^K(t_{-i})$ such that

$$u_i((a_i, a_{-i}), (t_i, t_{-i})) \geq u_i((a_i', a_{-i}), (t_i, t_{-i})).$$

Since $NSD_{-i}^K(t_{-i}) = NSD_{-i}(t_{-i})$, letting $t_j' = (t_j, a_j)$ for any $j \neq i$, we have $t_{-i}' \in B_i'(t_i')$ and

$$u_i'((\boldsymbol{s}_i(t_i'), \boldsymbol{s}_{-i}(t_{-i}')), (t_i', t_{-i}')) \geq u_i'((a_i', \boldsymbol{s}_{-i}(t_{-i}')), (t_i', t_{-i}')).$$

Thus player i is rational at type t_i', and $RAT_i = T'$ as desired. In sum, $\mathbf{CB}(RAT) = T'$ and Lemma S3 holds. ■

References

[1] D. Abreu and H. Matsushima. Virtual Implementation in Iteratively Undominated Actions: Incomplete Information. Working paper, 1992.

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

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

[4] P. Battigali and M. Siniscalchi. Rationalization and Incomplete Information. *Advances of Theoretical Economics*, Volume 3, Issue 1, Article 3, 46 pages, 2003.

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

[6] D. Bergemann and S. Morris. Informational Robustness and Solution Concepts. Working paper, 2014.

[7] B. Bernheim. Rationalizable Strategic Behavior. *Econometrica*, 52(4): 1007-1028, 1984.

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

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

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

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

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

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

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

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

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