The Japanese Economic Review
The Journal of the Japanese Economic Association

The Japanese Economic Review
Vol. 63, No. 1, March 2012

doi: 10.1111/j.1468-5876.2011.00553.x

ROBUST RATIONALIZABILITY UNDER ALMOST COMMON CERTAINTY OF PAYOFFS

By STEPHEN MORRIS†, SATORU TAKAHASHI† and OLIVIER TERCIEUX‡

†Princeton University ‡Paris School of Economics

An action is *robustly rationalizable* if it is rationalizable for every type who has almost common certainty of payoffs. We illustrate by means of an example that an action may not be robustly rationalizable even if it is weakly dominant, and argue that robust rationalizability is a very stringent refinement of rationalizability. Nonetheless, we show that every strictly rationalizable action is robustly rationalizable. We also investigate how permissive robust rationalizability becomes if we require that players be fully certain of their own payoffs.
JEL Classification Numbers: C72, D82.

1. Introduction

In his seminal paper, Rubinstein (1989) shows that equilibrium outcomes are highly sensitive to players' higher-order beliefs. In particular, he constructs a so-called "e-mail game" in which an action profile which is an equilibrium under common certainty of payoffs is not necessarily played in any equilibrium (or even in any interim approximate equilibrium) under "almost common certainty" in the sense of a high number of orders of certainty. That is, even if player i is certain about payoffs, i is certain that player j is certain about payoffs, i is certain that j is certain that i is certain about payoffs, and so on up to the n-th order, if i's higher-order beliefs are completely unrestricted, then he may not play actions that are equilibria under common certainty of payoffs. Weinstein and Yildiz (2007) extend this logic to rationalizability and show that only if an action is the unique rationalizable action for a type can one show that it is a rationalizable action for all nearby types in the product (i.e., pointwise convergence) topology over higher-order beliefs.

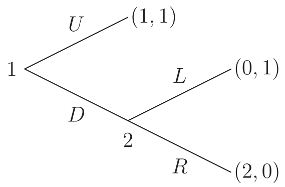
Monderer and Samet (1989), however, argue that the discontinuity of equilibrium actions identified by Rubinstein (1989) disappears if one uses "common p-belief" as an alternative notion of approximate common certainty. In particular, they show that if an action profile is an equilibrium under common certainty of payoffs, it is an interim approximate equilibrium (to any degree of precision) if there is common $(1 - \varepsilon)$-belief about payoffs for sufficiently small $\varepsilon > 0$. This result has a straightforward rationalizability counterpart: if an action is rationalizable under common certainty of payoffs, then it is approximately rationalizable (to any degree of precision) if there is common $(1 - \varepsilon)$-belief of payoffs for sufficiently small $\varepsilon > 0$. Here, we say that player i has common p-belief about an event if player i believes in the event with probability of at least p, player i believes with probability of at least p that player j believes in the event with probability of at least p, and so *ad infinitum*. Alternatively, one can regard common $(1 - \varepsilon)$-belief as (a variant of) the uniform convergence topology over higher-order beliefs. For example, Rubinstein's e-mail game is close to its complete-information limit in the product topology, but admits no type with common p-belief about payoffs with $p > 1/2$.[1,2]

In both Monderer and Samet's (1989) result and its rationalizability counterpart, the notion of rationality is relaxed to approximate rationality in incomplete-information games. Then the question arises: what happens to these results if we require exact rationality? Takahashi and Tercieux (2011) address this question for the equilibrium case. This paper reports some new results for the rationalizability case.

Consider an incomplete-information environment where players have uncertainty about payoff functions but maintain common certainty of full rationality, given their beliefs about those payoff functions. We say that an action is *robustly rationalizable* if it is rationalizable for *every* type who has common $(1 - \varepsilon)$-belief about payoffs for sufficiently small $\varepsilon > 0$.[3]

Under this definition, robustly rationalizable actions are always rationalizable, and conversely, every rationalizable action is robustly rationalizable in generic normal-form games. Robust rationalizability is, however, a very stringent refinement of rationalizability in sequential-move games, as illustrated in the following example.

Example 1: *Consider the following two-stage game*



and its normal-form representation

$$\theta^* : \begin{array}{c|c|c|} & L & R \\ \hline U & 1,\ 1 & 1,\ 1 \\ \hline D & 0,\ 1 & 2,\ 0 \\ \hline \end{array},$$

where θ^* is a payoff-relevant state that corresponds to this game. At state θ^*, L weakly dominates R. Also, in any complete-information game whose payoffs are sufficiently

[1] Kajii and Morris (1997) consider a somewhat intermediate case, where types are generated by some common prior and players are mutually certain of their own payoffs with high probability. In this case, they showed that with high probability, there is common p-belief about payoffs with $p < 1/|I|$, where $|I|$ is the number of players. Oyama and Tercieux (2010) discuss how much their result relies on the common prior assumption and in particular, how much common p-belief about payoffs must be weakened when players may have *ex ante* heterogeneous beliefs.

[2] One can ask the converse of Monderer and Samet's result: what is the weakest topology over higher-order beliefs that guarantees the continuity of equilibria or rationalizable actions? This question is addressed by Monderer and Samet (1996), Kajii and Morris (1998), Dekel *et al*. (2006), and Chen *et al*. (2010).

[3] By contrast, the literature has investigated various notions of robustness that test behavior against *some* sequence of nearby incomplete-information games. For example, Fudenberg *et al*. (1988) define the robustness of an equilibrium by checking if there *exists* a nearby incomplete-information game in which the equilibrium is played as a strict equilibrium. Dekel and Fudenberg (1990) ask a similar question but replace an equilibrium by a rationalizable action and a strict equilibrium by an action that survives iterated elimination of weakly dominated actions. Ely (2001) and Hu (2007) consider actions that can be rationalizable in *some* incomplete-information environment with almost common certainty of payoffs, and show that those actions are precisely rationalizable actions in the complete-information game.

close to those at state θ^*, U (respectively, L) is rationalizable for player 1 (respectively, player 2). (Whether D and R are rationalizable or not depends on payoff perturbations.)[4] Nevertheless, we argue that no action is robustly rationalizable in this game. For this purpose, we consider two more states:

$$\theta_{DL} : \begin{array}{c|c|c|} & L & R \\ \cline{2-3} U & 1,\ 1 & 1,\ 1 \\ \cline{2-3} D & \mathbf{2},\ 1 & 2,\ 0 \\ \cline{2-3} \end{array}, \qquad \theta_{DR} : \begin{array}{c|c|c|} & L & R \\ \cline{2-3} U & 1,\ 1 & 1,\ 1 \\ \cline{2-3} D & \mathbf{2},\ 1 & 2,\ \mathbf{2} \\ \cline{2-3} \end{array}.$$

Here boldface numbers represent payoffs that differ from corresponding payoffs in the original game.[5] In state θ_{DL}, D is the dominant strategy for player 1, and L is the strict best response to D for player 2 (as in state θ^*). In state θ_{DR}, D is the dominant strategy for player 1, and R is the strict best response to D for player 2. Player 1 has three types t_1, t_1', and t_1''; player 2 has two types t_2 and t_2'. Types and states are generated by the following common prior with arbitrarily small $\varepsilon > 0$ (a common prior is used for expository convenience only):

$$\begin{array}{c|c|c|} \theta^* & t_2 & t_2' \\ \cline{2-3} t_1 & 1/2 & 1/2 - 4\varepsilon \\ \cline{2-3} t_1' & 0 & \varepsilon \\ \cline{2-3} t_1'' & 0 & 0 \\ \cline{2-3} \end{array} \qquad \begin{array}{c|c|c|} \theta_{DL} & t_2 & t_2' \\ \cline{2-3} t_1 & 0 & 0 \\ \cline{2-3} t_1' & 0 & 0 \\ \cline{2-3} t_1'' & \varepsilon & 0 \\ \cline{2-3} \end{array} \qquad \begin{array}{c|c|c|} \theta_{DR} & t_2 & t_2' \\ \cline{2-3} t_1 & 0 & 0 \\ \cline{2-3} t_1' & 0 & 0 \\ \cline{2-3} t_1'' & 0 & 2\varepsilon \\ \cline{2-3} \end{array}$$

Note that all types except t_1'' have common p-belief with $p = (1 - 6\varepsilon)/(1 - 2\varepsilon) \approx 1$ that the state is θ^*. In this incomplete-information game, D is the unique rationalizable action for type t_1'' since he is certain that the state is either θ_{DL} or θ_{DR}. Given this, L is the unique rationalizable action for type t_2 since she weakly prefers L to R if the opponent's type is t_1, and her preference is strict if the opponent's type is t_1'', who plays D. Given this, U is the unique rationalizable action for type t_1 since he puts probability more than 1/2 that the opponent's type is t_2, who plays L. Then, R is the unique rationalizable action for type t_2' since she is indifferent between L and R if the opponent's type is t_1, who plays U, but conditional on that the opponent plays D, she puts probability of at least 2/3 that the opponent's type is t_1'', in which case the state is θ_{DR} and the opponent plays D. Finally, one can check that D is the unique rationalizable action for type t_1' since he is certain that the opponent's type is t_2', who plays R. Gathering all those arguments, we conclude that no action is robustly rationalizable in this game, and in particular, the uniqueness of perfect equilibrium is not sufficient for robust rationalizability.[6]

[4] If we see (U, L) as an action profile, it is not a strict equilibrium, but satisfies various equilibrium refinements. For example, $\{(U, L)\}$ is a persistent retract, hence (U, L) is strictly perfect and proper. Moreover (U, L) is essential (van Damme, 1991).

[5] Note that payoffs from (U, L) and (U, R) are identical in all states, and hence one can see those states as states that affect payoffs on terminal nodes in the extensive form without changing the game tree.

[6] Note that type t_1 occurs with prior probability ε only. Thus, if one relaxed the definition of robustness and required that an action be played with high prior probability, one could argue that U is robustly rationalizable in this weaker sense. Similar definitions are adopted in Kajii and Morris (1997) and Takahashi and Tercieux (2011). We do not, however, take this route in this paper, for we are afraid of introducing *ex ante* perspectives in a setup with otherwise exclusively interim perspectives. See Oyama and Tercieux (2010), who discuss the relation between robustness notions from *ex ante* and interim perspectives.

The goal of this paper is to provide a sufficient condition for robust rationalizability. We say that an action is *strictly rationalizable* if it survives iterated elimination of actions that are not strict best responses to any conjecture over the remaining actions of the opponents. We show that every strictly rationalizable action is robustly rationalizable.

We also investigate how the notion of robust rationalizability is affected if we require that players be fully certain of their own payoffs. That is, we say that an action is *C-robustly rationalizable* (where C stands for Certainty) if it is rationalizable for every type who has common $(1 - \varepsilon)$-belief about certainty of own payoffs for sufficiently small $\varepsilon > 0$. We show that unlike the original definition of robust rationalizability, the weakly dominant action, if it exists, is C-robustly rationalizable. More generally, a slight weakening of strict rationalizability is sufficient for C-robust rationalizability.

To conclude the introduction, we briefly mention related work on refinements of rationalizability, such as perfect rationalizability in Bernheim (1984), cautious rationalizability in Pearce (1984), $\mathbf{S}^{\infty}\mathbf{W}$ in Dekel and Fudenberg (1990) and Börgers (1994), proper rationalizability in Schuhmacher (1999), and weakly perfect and trembling-hand perfect rationalizability in Herings and Vannetelbosch (1999, 2000). There are some technical differences between those concepts, but they all share the property that weakly dominated actions are eliminated. Unlike our notion of (C-)robust rationalizability, for each of such refinements, the set of rationalizable actions that pass the refinement is non-empty in every game. Also, among those papers, it is only Dekel and Fudenberg (1990) who explicitly formalize refinements as a question of rationalizability in incomplete-information games.

2. Definitions

2.1 Complete-information games

We consider a complete-information game $G = (I\ (A_i, g_i)_{i\in I})$, where I is the set of players, and for each $i \in I$, A_i is the set of actions available to player i, $A = \prod_i A_i$, and $g_i : A \to \mathbb{R}$ is his payoff function. We assume that both I and A are finite. The domain of g_i is extended to $A_i \times \Delta(A_{-i})$ in the usual way, where $A_{-i} = \prod_{j\neq i} A_j$ and for each measurable space X, $\Delta(X)$ denotes the set of probability measures over X. For each (correlated) probability $\lambda_i \in \Delta(A_{-i})$, let

$$br_i(\lambda_i) = \operatorname*{argmax}_{a_i\in A_i} \sum_{a_{-i}\in A_{-i}} \lambda_i(a_{-i})\, g_i(a_i, a_{-i})$$

be the set of best responses against λ_i for player i.

We define rationalizability by iteratively eliminating actions that are not best responses to any conjecture about the remaining actions of the opponents. That is, let $R_i^0 = A_i$ for each $i \in I$, and for each $n \geq 1$, let R_i^n be the set of best responses against $R_{-i}^{n-1} = \prod_{j\neq i} R_j^{n-1}$, that is, the set of actions $a_i \in A_i$ that satisfy $a_i \in br_i(\lambda_i)$ for some $\lambda_i \in \Delta\left(R_{-i}^{n-1}\right)$. Then $\left\{R_i^n\right\}$ is a decreasing sequence (with respect to the set-inclusion order). Let $R_i = \bigcap_{n\geq 0} R_i^n$, and we say that any $a_i \in R_i$ is (correlated) *rationalizable*.[7] Note that the profile $(R_i)_{i\in I}$ of the sets of rationalizable actions is characterized by the largest profile $(X_i)_{i\in I}$ of sets that satisfies the following fixed-point property: for each $i \in I$, $X_i \subseteq A_i$ is equal to (or a subset of) the set of all best responses against $X_{-i} = \prod_{j\neq i} X_j \subseteq A_{-i}$. Also note that R_i is non-empty.

2.2 Incomplete-information elaborations

We embed G into an incomplete-information environment. Let Θ be a finite set of payoff-relevant states, where each player $i \in I$ has a state-dependent payoff function $u_i : A\times\Theta \to \mathbb{R}$. We assume that Θ contains a state θ^* such that $u_i(\cdot, \theta^*) = g_i$ for all $i \in I$. Let $\mathcal{T} = (T_i, \pi_i)_{i\in I}$ be the universal type space over Θ. Here, T_i is the set of universal types, each of which specifies a hierarchy of beliefs, i.e., player i's belief over Θ, his belief about his opponents' beliefs about Θ, his belief about his opponents' beliefs about his opponents' opponents' beliefs about Θ, etc., and π_i: $T_i \to \Delta(T_{-i} \times \Theta)$ is the natural homeomorphism.[8] Following Monderer and Samet (1989), for $p \in [0, 1]$, we define p-belief operators and common p-belief as follows. For each product event $E_{-i}\times\Psi = \prod_{j\neq i} E_j \times \Psi \subseteq T_{-i}\times\Theta$, let $B_i^p(E_{-i}\times\Psi)$ be the set of player i's types who puts probability of at least p on $E_{-i} \times \Psi$:

$$B_i^p(E_{-i}\times\Psi) = \{t_i \in T_i | \pi_i(t_i)(E_{-i}\times\Psi) \geq p\}.$$

For each product event $E\times\Psi = \prod_i E_i \times\Psi \subseteq T\times\Theta$, we define the event where there is common p-belief about $E \times \Psi$ by iteratively applying p-belief operators B_i^p to $E \times \Psi$. That is, let $E_i^0 = E_i$ for each $i \in I$, and for each $n \geq 1$, let $E_i^n = B_i^p\left(E_{-i}^{n-1}\times\Psi\right)\cap E_i$. Let $CB_i^p(E\times\Psi) = \bigcap_{n\geq 0} E_i^n$, and we say that any $t_i \in CB_i^p(E\times\Psi)$ is in E_i and has common p-belief about $E \times \Psi$. We have $B_i^p\left(CB_{-i}^p(E\times\Psi)\times\Psi\right)\cap E_i = CB_i^p(E\times\Psi)$.

Given (I, A, Θ, u), we can define an incomplete-information game on the universal type space $\mathcal{T}$, where I is the set of players, T_i is the set of types of player i, A_i is the set of actions available to any type in T_i, u_i is his state-dependent payoff function, and π_i describes his belief over the opponents' types and states. We call this game an *elaboration of G*. In this game, we define interim correlated rationalizablity introduced by Dekel *et al.* (2007). The definition goes as follows. Let $ICR_i^0(t_i) = A_i$ for each $i \in I$ and $t_i \in T_i$. For each $n \geq 1$, let $ICR_i^n(t_i)$ be the set of best responses for player i against some beliefs over his opponents' types and actions and the state that put probability 1 on the opponents playing actions within ICR_{-i}^{n-1} and are consistent with $\pi_i(t_i)$, that is,

$$ICR_i^n(t_i) = \left\{ a_i \in A_i \,\middle|\, \begin{array}{l} \exists\mu_i \in \Delta(T_{-i}\times A_{-i}\times\Theta)\text{ s.t.} \\ \left(\operatorname{marg}_{T_{-i}\times A_{-i}}\mu_i\right)\left(\text{graph } ICR_{-i}^{n-1}\right) = 1, \\ \operatorname{marg}_{T_{-i}\times\Theta}\mu_i = \pi_i(t_i), \\ \forall a_i' \in A_i, u_i\left(a_i, \operatorname{marg}_{A_{-i}\times\Theta}\mu_i\right) \geq u_i\left(a_i', \operatorname{marg}_{A_{-i}\times\Theta}\mu_i\right) \end{array} \right\},$$

[7] In games with more than two players, correlated rationalizability is more permissive than independent rationalizability as originally defined by Bernheim (1984) and Pearce (1984), which requires that each player has a stochastically independent conjecture about the opponents' actions.

[8] See Mertens and Zamir (1985) and Brandenburger and Dekel (1993) for constructions of the universal type space.

where

$$\text{graph } ICR_{-i}^{n-1} = \{(t_{-i}, a_{-i}) \in T_{-i} \times A_{-i} | \forall j \neq i, a_j \in ICR_j^{n-1}(t_j)\},$$

and the domain of u_i is extended to $A_i \times \Delta(A_{-i} \times \Theta)$ in the usual way. Let $ICR_i(t_i) = \bigcap_{n \geq 0} ICR_i^n(t_i)$, and we say that any $a_i \in ICR_i(t_i)$ is *interim correlated rationalizable* (*ICR*) for type t_i. Note that each ICR_i is not a set of actions, but a correspondence that maps each type to a set of actions. Also note that in each step of iteration of ICR_i^n, each type has a belief over the opponents' types, actions, and the state such that the action a_j of one of his opponents can be correlated with other opponents' actions and types as well as the state. Thus ICR is a very (probably the most) permissive interim solution concept defined in incomplete-information games that allow for hidden correlation devices across players and between players and the nature. Dekel *et al*. (2007) show that ICR is invariant with respect to belief-preserving morphisms and thus defined without loss of generality in the universal type space. They also show that ICR admits a simple epistemic characterization in terms of common certainty of rationality.

3. Robust rationalizability

We say that an action is robustly rationalizable in a complete-information game if it is played in ICR for every type with almost common certainty of payoffs.

Definition 1: *Fix a complete-information game G. An action $a_i \in A_i$ is* robustly rationalizable *for player* i *in* G *if, for any pair* (Θ, u) *of payoff-relevant states and state-dependent payoff functions, there exists* $\varepsilon > 0$ *such that for any $t_i \in CB_i^{1-\varepsilon}(T \times \{\theta^*\})$, we have $a_i \in ICR_i(t_i)$.*

Since $ICR_i(t_i) = R_i$ for $t_i \in CB^1(T \times \{\theta^*\})$, rationalizability is necessary for robust rationalizability. Here we explore a sufficient condition for robust rationalizability.

For this purpose, we define strict rationalizability in a complete-information game G. Strict rationalizability is similar to rationalizability, but in each step of iteration, we eliminate actions that are not strict best responses to any conjecture about the remaining actions of the opponents. That is, let $R_i^{s,0} = A_i$ for each $i \in I$, and for each $n \geq 1$, let $R_i^{s,n}$ be the set of strict best responses against $R_{-i}^{s,n-1}$, that is, the set of actions $a_i \in A_i$ that satisfy $\{a_i\} = br_i(\lambda_i)$ for some $\lambda_i \in \Delta(R_{-i}^{s,n-1})$. Let $R_i^s = \bigcap_{n \geq 0} R_i^{s,n}$, and we say that any $a_i \in R_i^s$ is *strictly rationalizable*. For each $i \in I$, R_i^s is equal to the set of all strict best responses against R_{-i}^s. We have $R_i^s \subseteq R_i$ in all games, and $R_i = R_i^s$ in generic normal-form games, but R_i^s may be the empty set in non-generic cases. Also, if a strict equilibrium $a^* = (a_i^*) \in A$ exists, then $a_i^* \in R_i^s$ for any $i \in I$.

Proposition 1: *Any strictly rationalizable action is robustly rationalizable.*

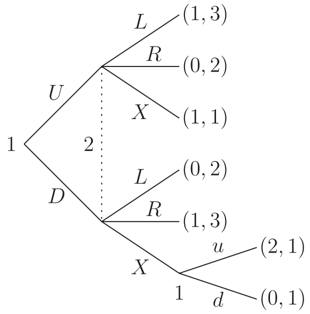
Proof: By the fixed-point property of (R_i^s), there exists $\varepsilon > 0$ such that for any $i \in I$ and any $a_i \in R_i^s$, there exists $\lambda_i^{a_i} \in \Delta(R_{-i}^s)$ such that

$$(1-\varepsilon)\left(g_i(a_i, \lambda_i^{a_i}) - g_i(a_i', \lambda_i^{a_i})\right) - \varepsilon \max_{a_{-i}, \theta}(u_i(a_i, a_{-i}, \theta) - u_i(a_i', a_{-i}, \theta)) \geq 0$$

for any $a_i' \in A_i$. Define the sequence $\{ E_i^n \}$ of sets of types who have n-th level of mutual $(1 - \varepsilon)$-belief about θ^* as follows. Let $E_i^0 = T_i$ for each $i \in I$, and for each $n \geq 1$, let $E_i^n = B_i^{1-\varepsilon}(E_{-i}^{n-1} \times \{\theta^*\})$. We show by induction on n that for any $i \in I$, if $t_i \in E_i^n$, then $ICR_i^n(t_i) \supseteq R_i^s$. The case with $n = 0$ is obvious. For $n \geq 1$, suppose that for any $i \in I$, if $t_i \in E_i^{n-1}$, then $ICR_i^{n-1}(t_i) \supseteq R_i^s$. Pick any $a_i \in R_i^s$ and any $t_i \in E_i^n = B_i^{1-\varepsilon}(E_{-i}^{n-1} \times \{\theta^*\})$. Let $\mu_i \in \Delta(T_{-i} \times A_{-i} \times \Theta)$ be a probability measure such that the marginal $\mathrm{marg}_{T_{-i}\times\Theta}\mu_i$ is given by $\pi_i(t_i)$, and conditional on t_{-i} and θ, action profile a_{-i} is chosen according to $\lambda_i^{a_i}$ if $t_{-i} \in E_{-i}^{n-1}$ and $\theta = \theta^*$, while chosen arbitrarily from $ICR_{-i}^{n-1}(t_{-i})$ if $t_{-i} \notin E_{-i}^{n-1}$ or $\theta \neq \theta^*$. Then, by the choice of ε, a_i is a best response against $\mathrm{marg}_{A_{-i}\times\Theta}\mu_i$. Thus $ICR_i^n(t_i) \supseteq R_i^s$. ■

In Example 1 in the Introduction, we have $R_1^s = \emptyset \subsetneq \{U, D\} = R_1$ and $R_2^s = \emptyset \subsetneq \{L, R\} = R_2$. Also, no action is robustly rationalizable.

Example 2: *Strict rationalizability is not necessary for robust rationalizability. To see this, consider the following extensive-form game*



and its reduced-normal-form representation

	L	R	X
U	1, 3	0, 2	1, 1
Du	0, 2	1, 3	2, 1
Dd	0, 2	1, 3	0, 1

(The following argument is insensitive to small payoff perturbations on terminal nodes in the extensive form.) In this game, we have $R_1^s = \{U\}$ and $R_2^s = \{L\}$. To see this, first note that X is strictly dominated by L. Once X is eliminated, actions Du and Dd are equivalent, hence neither Du nor Dd can be a strict best response to any conjecture over $\{L, R\}$. Thus both Du and Dd are eliminated. Then R is eliminated since L is the strict best response against U. We can show, however, that R is robustly rationalizable for player 2. More strongly, we can show that for any t_1 and t_2 with almost common certainty of θ^*, $ICR_1(t_1)$ is either $\{U, Du, Dd\}$, $\{U, Du\}$, or $\{U, Dd\}$, i.e., $ICR_1(t_1)$ contains U as well as either Du or Dd (or both), whereas $ICR_2(t_2) = \{L, R\}$. This is because type t_1 can rationalize either Du or Dd by conjecturing that player 2 plays R with high probability, while type t_2 can rationalize R by conjecturing that player 1 plays Du or Dd with high probability, without specifying which one to be played.

4. Certainty of own payoffs

In Example 1, for type t_2' to eliminate the weakly dominant action L in the original game, it was essential that she is *not* certain of her own payoffs. Instead, she puts a small but positive probability on the event that player 1's type is t_1'', and conditional on that event, R is no longer weakly dominated in her payoff function. In this section, we explore the implications of assuming that players are fully certain of their own payoffs.

4.1 Almost common certainty of certainty of own payoffs

We define robustness of rationalizable actions with respect to types who are certain of their own payoffs. Let E_i^* be the set of player i's type who is certain that his own payoffs are given by g_i:

$$E_i^* = B_i^1(T_{-i} \times \{\theta \in \Theta | u_i(\cdot,\theta) = g_i\})$$

$$= \{t_i \in T_i | \pi_i(t_i)(T_{-i} \times \{\theta \in \Theta | u_i(\cdot,\theta) = g_i\}) = 1\}.$$

Definition 2: *An action $a_i \in A_i$ is* C-robustly rationalizable *for player* i *in* G *if, for any pair* (Θ, u) *of payoff-relevant states and state-dependent payoff functions, there exists $\varepsilon > 0$ such that for any $t_i \in CB_i^{1-\varepsilon}(E^* \times \Theta)$, we have $a_i \in ICR_i(t_i)$.*

In a complete-information game G, we say that a_i is a strictly perfect best response against X_{-i} if there exists $\varepsilon > 0$ such that for any $\mu_i \in \Delta(A_{-i})$, there exists $\lambda_i \in \Delta(X_{-i})$ such that $a_i \in br_i((1 - \varepsilon)\lambda_i + \varepsilon\mu_i)$. Let $R_i^{\mathrm{sp},0} = A_i$, and for each $n \geq 1$, let $R_i^{\mathrm{sp},n}$ be the set of strictly perfect best responses against $R_{-i}^{\mathrm{sp},n-1}$. Let $R_i^{\mathrm{sp}} = \bigcap_{n\geq 0} R_i^{\mathrm{sp},n}$, and we say that any $a_i \in R_i^{\mathrm{sp}}$ is *strictly perfectly rationalizable*.[9] Note that R_i^1 is the set of admissible (i.e., undominated) actions against $R_{-i}^0 = A_{-i}$, but the construction of R_i^n for $n \geq 2$ is more involved. We have $R_i^{\mathrm{s}} \subseteq R_i^{\mathrm{sp}} \subseteq R_i$, and any weakly dominant action is strictly perfectly rationalizable.

Proposition 2: *Any strictly perfectly rationalizable action is C-robustly rationalizable.*

Proof: By the fixed-point property of R_i^{sp}, there exists $\varepsilon > 0$ such that for any $i \in I$, $a_i \in R_i^{\mathrm{sp}}$, and for $\mu_i \in \Delta(A_{-i})$, there exists $\lambda_i^{a_i,\mu_i} \in \Delta\left(R_{-i}^{\mathrm{sp}}\right)$ such that

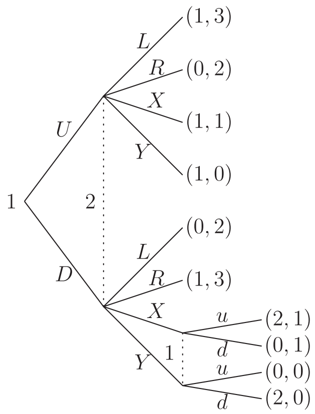
$$a_i \in br_i\left((1-\varepsilon)\lambda_i^{a_i,\mu_i} + \varepsilon\mu_i\right).$$

The rest is similar to the proof of Proposition 1. ■

In Example 1, $R_1^{\mathrm{sp}} = \{U\}$ and $R_2^{\mathrm{sp}} = \{L\}$. It is easy to see that U (respectively, L) is the unique C-robustly rationalizable action for player 1 (respectively, player 2).

[9] Strictly perfect rationalizability is a rationalizability version of strictly perfect equilibria in Okada (1981).

Example 3: *Strictly perfect rationalizability is not necessary for C-robust rationalizability. To see this, consider the following extensive-form game*



and its reduced-normal-form representation

	L	*R*	*X*	*Y*
U	1, 3	0, 2	1, 1	1, 0
Du	0, 2	1, 3	2, 1	0, 0
Dd	0, 2	1, 3	0, 1	2, 0

Since X and Y are strictly dominated, we have $R_2^{\mathrm{sp}} \subseteq \{L, R\}$. Then Du is never a best response to $(1-\varepsilon)\lambda_i + \varepsilon Y$ for any $\lambda_i \in \Delta(\{L, R\})$. Similarly, Dd is never a best response to $(1-\varepsilon)\lambda_i + \varepsilon X$ for any $\lambda_i \in \Delta(\{L, R\})$. Thus $R_1^{\mathrm{sp}} = \{U\}$ and $R_2^{\mathrm{sp}} = \{L\}$. We can show, however, that R is C-robustly rationalizable for player 2. To see this, by the same argument as in Example 2, for any t_1 and t_2 with almost common certainty of certainty of own payoffs, $ICR_1(t_1)$ is either $\{U, Du, Dd\}$, $\{U, Du\}$, or $\{U, Dd\}$, and $ICR_2(t_2) = \{L, R\}$. Type t_1 can rationalize either Du or Dd by conjecturing that player 2 plays R with high probability, while type t_2 can rationalize R by conjecturing that player 1 plays Du or Dd with high probability.

4.2 Common *p*-belief about certainty of own payoffs

We can generalize the previous subsection and investigate, for a fixed value of $p \in (0,1)$, the robustness of rationalizability when there is common p-belief about certainty of own payoffs.

Definition 3: *For* p ∈ *(0, 1), an action* $a_i \in A_i$ *is* p-robustly rationalizable *for player* i *in* G *if, for any pair* (Θ, u) *of payoff-relevant states and state-dependent payoff functions and any* $t_i \in CB_i^p(E^* \times \Theta)$, *we have* $a_i \in ICR_i(t_i)$.

The definition of p-perfect rationalizability mimics the definition of strictly perfect rationalizability but fixes $\varepsilon = 1 - p$. More precisely, we say that a_i is p-perfect best response against X_{-i} if for any $\mu_i \in \Delta(A_{-i})$, there exists $\lambda_i \in \Delta(X_{-i})$ such that $a_i \in br_i(p\lambda_i + (1-\mathrm{p})\mu_i)$. Let $R_i^{p,0} = A_i$, and for each $n \geq 1$, let $R_i^{p,n}$ be the set of p-perfect best responses against $R_{-i}^{p,n-1}$. Let $R_i^p = \bigcap_{n \geq 0} R_i^{p,n}$, and we say that any $a_i \in R_i^p$ is *p-perfectly rationalizable.*[10] R_i^p is weakly increasing in the set-inclusion order and converges to R_i^{sp} as $p \to 1$, i.e., for $0 < p < q < 1$, we have $R_i^p \subseteq R_i^q \subseteq R_i^{\mathrm{sp}}$ and $\bigcup_{p<1} R_i^p = R_i^{\mathrm{sp}}$. Also, any weakly dominant action is p-perfectly rationalizable for any p.

The following is a simple extension of Proposition 2 and hence the proof is omitted.

Proposition 3: *Any p-perfectly rationalizable action is p-robustly rationalizable.*

5. Concluding remarks

We introduced the notion of robust rationalizability with respect to small uncertainty about payoffs. We then provided various sufficient conditions for an action to be robustly rationalizable, depending on which class of incomplete-information perturbations we use to test the robustness. As illustrated in a series of examples, none of these conditions is, however, a full characterization for (C-)robust rationalizability. It is left for future research to explore such if-and-only-if results.

As we pointed out in the Introduction, our analysis hinges crucially on the requirement of full rationality in the definition of ICR. If instead we use approximate rationality and define the set $ICR_i^\delta(t_i)$ of δ-ICR actions for type t_i by relaxing the incentive constraint to

$$u_i\left(a_i, \operatorname{marg}_{A_{-i}\times\Theta}\mu_i\right) \geq u_i\left(a_i', \operatorname{marg}_{A_{-i}\times\Theta}\mu_i\right) - \delta,$$

then, one can show that any rationalizable action $a_i \in R_i$ is robustly rationalizable in the sense that, for any (Θ, u) and $\delta > 0$, there exists $\varepsilon > 0$ such that for any $t_i \in CB_i^{1-\varepsilon}(T \times \{\theta^*\})$, we have $a_i \in ICR_i^\delta(t_i)$ (Dekel *et al*., 2006). Implications of those differences between exact ICR and δ-ICR are also explored in Chen and Xiong (2011) under the product topology on the universal type space.

Final version accepted 11 October 2011.

References

Bernheim, D. (1984) "Rationalizable Strategic Behavior", *Econometrica*, Vol. 52, No. 4, pp. 1007–1028.

Börgers, T. (1994) "Weak Dominance and Approximate Common Knowledge", *Journal of Economic Theory*, Vol. 64, No. 1, pp. 265–276.

Brandenburger, A. and E. Dekel (1993) "Hierarchies of Beliefs and Common Knowledge", *Journal of Economic Theory*, Vol. 59, No. 1, pp. 189–198.

Chen, Y.-C. and S. Xiong (2011) "Robust Selection of Rationalizability", working paper.

——, A. Di Tillio, E. Faingold and S. Xiong (2010) "Uniform Topologies on Types", *Theoretical Economics*, Vol. 5, No. 3, pp. 445–478.

van Damme, E. (1991) *Stability and Perfection of Nash Equilibria*, Berlin: Springer-Verlag.

Dekel, E. and D. Fudenberg (1990) "Rational Behavior under Payoff Uncertainty", *Journal of Economic Theory*, Vol. 52, No. 2, pp. 243–267.

——, —— and S. Morris (2006) "Topologies on Types", *Theoretical Economics*, Vol. 1, No. 3, pp. 275–309.

——, —— and —— (2007) "Interim Correlated Rationalizability", *Theoretical Economics*, Vol. 2, No. 1, pp. 15–40.

[10] Note the difference from p-rationalizability in Hu (2007), where the set $\tilde{R}_i^p = \bigcap_{n\geq 0} \tilde{R}_i^{p,n}$ of p-rationalizable actions is defined by $\tilde{R}_i^{p,0} = A_i$ and for any $n \geq 1$, $\tilde{R}_i^{p,n}$ is the set of best responses against $p\lambda_i + (1-p)\mu_i$ for some $\lambda_i \in \Delta\left(\tilde{R}_{-i}^{p,n-1}\right)$ and *some* $\mu_i \in \Delta(A_{-i})$.

Ely, J. (2001) “Rationalizability and Approximate Common-Knowledge”, working paper.

Fudenberg, D., D. Kreps and D. Levine (1988) “On the Robustness of Equilibrium Refinements”, *Journal of Economic Theory*, Vol. 44, No. 2, pp. 354–380.

Herings, J. J. and V. J. Vannetelbosch (1999) “Refinements of Rationalizability for Normal-Form Games”, *International Journal of Game Theory*, Vol. 28, No. 1, pp. 53–68.

—— and —— (2000) “The Equivalence of the Dekel-Fudenberg Iterative Procedure and Weakly Perfect Rationalizability”, *Economic Theory*, Vol. 15, No. 3, pp. 677–687.

Hu, T.-W. (2007) “On *P*-Rationalizability and Approximate Common Certainty of Rationality”, *Journal of Economic Theory*, Vol. 136, No. 1, pp. 379–391.

Kajii, A. and S. Morris (1997) “The Robustness of Equilibria to Incomplete Information”, *Econometrica*, Vol. 65, No. 6, pp. 1283–1309.

—— and —— (1998) “Payoff Continuity in Incomplete Information Games”, *Journal of Economic Theory*, Vol. 82, No. 1, pp. 267–276.

Mertens, J.-F. and S. Zamir (1985) “Formulation of Bayesian Analysis for Games with Incomplete Information”, *International Journal of Game Theory*, Vol. 14, No. 1, pp. 1–29.

Monderer, D. and D. Samet (1989) “Approximating Common Knowledge with Common Beliefs”, *Games and Economic Behavior*, Vol. 1, No. 2, pp. 170–190.

—— and —— (1996) “Proximity of Information in Games with Incomplete Information”, *Mathematics of Operations Research*, Vol. 21, No. 3, pp. 707–725.

Okada, A. (1981) “On Stability of Perfect Equilibrium Point”, *International Journal of Game Theory*, Vol. 10, No. 2, pp. 67–73.

Oyama, D. and O. Tercieux (2010) “Robust Equilibria under Non-Common Priors”, *Journal of Economic Theory*, Vol. 145, No. 2, pp. 752–784.

Pearce, D. (1984) “Rationalizable Strategic Behavior and the Problem of Perfection”, *Econometrica*, Vol. 52, No. 4, pp. 1029–1050.

Rubinstein, A. (1989) “The Electronic Mail Game: A Game with Almost Common Knowledge”, *American Economic Review*, Vol. 79, No. 3, pp. 389–391.

Schuhmacher, F. (1999) “Proper Rationalizability and Backward Induction”, *International Journal of Game Theory*, Vol. 28, No. 4, pp. 599–615.

Takahashi, S. and O. Tercieux (2011) “Robust Equilibria in Sequential Games under Almost Common Certainty”, working paper.

Weinstein, J. and M. Yildiz (2007) “A Structure Theorem for Rationalizability with Application to Robust Predictions of Refinements”, *Econometrica*, Vol. 75, No. 2, pp. 365–400.