HIER

Harvard Institute of Economic Research

Discussion Paper Number 2064

Interim Rationalizability

by

Eddie Dekel, Drew Fudenberg
and Stephen Morris

March 2005

Harvard University
Cambridge, Massachusetts

This paper can be downloaded without charge from:
http://post.economics.harvard.edu/hier/2005papers/2005list.html

The Social Science Research Network Electronic Paper Collection:
http://ssrn.com/abstract=676672

Abstract

This paper proposes the solution concept of *interim rationalizability*, and shows that all type spaces that have the same hierarchies of beliefs have the same set of interim rationalizable outcomes. This solution concept characterizes common knowledge of rationality in the universal type space. JEL Classification and keywords: C70,C72, rationalizability, incomplete information, common knowledge, universal type space.

1 Introduction

Harsanyi (1967-8) proposes solving games of incomplete information using type spaces, and Mertens and Zamir (1985) show how to construct a universal type space, into which all other type spaces (satisfying certain technical regularity assumptions) can be mapped.[1] The universal type space is the set of all infinite hierarchies of beliefs satisfying common knowledge of coherency. However, Bergemann and Morris (2001, section 2.2.2) and Battigalli and Siniscalchi (2003) emphasize that type spaces may allow for more correlation than is captured in the belief hierarchies, so identifying types that have identical hierarchies may lead to a loss of information, and solution concepts can differ when applied to two different type spaces even if they are mapped into the same subset of the universal type space. This paper proposes the solution concept of interim correlated rationalizability, and in Proposition 2 shows that all type spaces that have the same hierarchies of beliefs have the same set of interim correlated rationalizable outcomes. Thus this is a solution concept which can be characterized by working with the universal type space, and more generally permits identifying those type spaces that have the same hierarchies of beliefs.

Brandenburger and Dekel (1987) showed that the set of actions that survive iterated deletion of strictly dominated strategies in a complete information game is equal to the set of actions that could be played in a subjective correlated equilibrium. Our second result, proposition 3, reports a straightforward extension of Brandenburger and Dekel's observation to games with incomplete information; this shows that interim rationalizability characterizes common knowledge of rationalizability.

We now sketch the main constructs in the paper. Fix a type space, where players have beliefs and higher order beliefs about some payoff relevant state space Θ. A game consists of payoff functions mapping from action profiles and Θ to the real line. We discuss two definitions of interim rationalizability. To find the set of independently interim rationalizable strategies, iteratively delete for each type all actions that are not best responses given that type's beliefs over others' types and Θ

[1] Further details and references are provided in the next section.

and given any profile of strategies of all other players, where under those strategies, each type puts positive probability only on surviving actions. The set of (correlated) interim rationalizable actions results from iteratively deleting for each type all actions that are not best responses given that type's beliefs over others' types and Θ and given any (perhaps correlated) conditional beliefs about which surviving actions are played by at a given type profile and payoff relevant state. In this definition, a player's beliefs allow for correlation between one player's actions and the payoff state and other players' actions. In the complete information case (i.e., when Θ is a singleton), these definitions reduce to the standard definitions of independent and (correlated) rationalizability, respectively (e.g., as in Brandenburger and Dekel (1987)). In the complete-information case, independent and correlated rationalizability are equivalent when there are two players but not necessarily with three or more players. We will see that with incomplete information, they may differ even in the two-person case (because of the possible correlation in a player's conjecture between the opponent's actions and the payoff-relevant state).

Our contribution in this paper is to clarify and slightly extend existing work. We use the concept of interim rationalizability discussed in this paper in our ongoing work on defining strategic topologies on the universal type space (Dekel, Fudenberg and Morris (2004)).[2] For this exercise, it is important for us to know that the solution concept depends only on hierarchies of beliefs and not on "redundant" elements of the type space. Another contribution of this note is to identify and analyze the distinction between independent and correlated interim rationalizability.

Battigalli and Siniscalchi (2003) define an umbrella notion of "Δ-rationalizable" actions in incomplete information environments, where Δ can be varied to capture common-knowledge restrictions on the first order of beliefs in the hierarchy. They show that there is an equivalence between actions surviving an iterative procedure capturing common knowledge of Δ and the set of actions that might be played in equilibrium on any type space where Δ is common knowledge. (Correlated) interim rationalizable actions are exactly Δ-rationalizable actions, where we let Δ consist of a complete description of the infinite hierarchies of beliefs. With this Δ, proposition 3 below corresponds to their proposition 4.3; they do not discuss explicitly mention either correlated or independent interim rationalizability.[3]

Forges (1993) explores the related question of how to define correlated equilibrium for games of incomplete information. Forges allows correlating devices that enable a player's own actions to

[2] For this purpose, we also verify below that all the results extend to ε-rationalizability and ε-equilibrium.

[3] As their analysis deals with restrictions on first order beliefs, our result corresponds to an extension of their approach to allow for restrictions on the entire hierarchy of beliefs.

depend on the payoff states θ even when the player cannot distinguish between the states; as we discuss at the end of subsection 3.4 this is similar to what we do. Furthermore, like Battigalli and Siniscalchi (2003), Forges studies the solution concept corresponding to allowing the type space to vary over all spaces (with a common prior). Forges' proposition 3, that relates common knowledge of rationality (with a common prior and a given type space) to agent-normal form correlated equilibrium is analogous to our remark 1 in which we reinterpret proposition 3 as implying that common knowledge of rationality corresponds to interim (correlated) rationalizability.

Weinstein and Yildiz (2003) prove that under additional assumptions (generic payoffs and a richness condition on the payoff uncertainty) the conclusion of our proposition 3 can be strengthened in certain directions. Pick any type t_i in the universal type space, and any equilibrium s of a game played on the universal type space, and any action a_i that is interim rationalizable for t_i. Then for any k, there is a type t_i' that agrees with t_i up to the first k steps in the hierarchy and such that $s_i(t_i) = a_i$. This result would follow from our Propsotion 3, if one were to add redundant types to the universal type space and to pick the equilibrium on the enlarged type space. But Weinstein and Yildiz show that—under their additional assumptions—any equilibrium on the universal type space itself already contains enough richness for the conclusion.

A recent paper by Ely and Peski (2004) also notes that the set of *independent* interim rationalizable outcomes depend on more than just the standard universal type space. They characterize how the standard universal type space must be expanded to deal with this issue. Thus, while we find the solution concept which depends on types only via their (standard) hierarchies of beliefs, Ely and Peski provide an extended notion of hierarchies of beliefs for which a different solution concept depends on types only via those extended hierarchies.

2 Type Spaces

We base our development of type spaces on Heifetz and Samet's (1998) topology-free construction. The primitive of the model is a measurable set of states of Nature, Θ, a finite set of players, I, and a type space $\mathcal{T} = (T_i, \widehat{\pi}_i)_{i=1}^{I}$, where each T_i is a measurable space and each $\widehat{\pi}_i : T_i \to \Delta(T_{-i} \times \Theta)$, where for measurable X we denote by $\Delta(X)$ the set of (probability) measures on X, and $T_{-i} \times \Theta$ is endowed with the product σ-algebra. Points $t_i \in T_i$ are called player *i's types*, and we say that each type t_i of player i has belief $\widehat{\pi}_i(t_i)$ about the joint distribution of the opponent's type and the state of Nature.

The set of measures, $\Delta(X)$, is endowed with the σ-field generated by

$$\{\{\mu : \mu(E) \geq p\} : p \in [0,1] \text{ and } E \subset X\}.$$

(Throughout, when writing $E \subset X$ for a measurable space X we consider only measurable subsets, and all functions are measurable; we do not specify this restriction again.)[4]

A type space can be translated into another type space if it can be mapped into that space while preserving the belief structure. Formally, (T_i, π_i) is translated into $\left(\tilde{T}_i, \tilde{\pi}_i\right)$ if for each i there exists $\varphi_i : T_i \to \tilde{T}_i$ with $\tilde{\pi}_i(\varphi_i(t_i))(F) = \pi_i(t_i)\left(\left\{(s, t_{-i}) : \left(s, \varphi_{-i}(t_{-i})\right) \in F\right\}\right)$ for $F \subset \Theta \times \tilde{T}_{-i}$.

A particular type space is Harsanyi's "universal type space," as constructed by Mertens and Zamir (1985).[5] Specifically, let $X_0 = \Theta$, and define $X_k = X_{k-1} \times [\Delta(X_{k-1})]^{I-1}$, where $\Delta(X_k)$ is the set of probability measures on the Borel field $B(X_k)$ of X_k, endowed with the "weak" topology, and each X_k is given the product topology over its two components. An element $(\delta_1, \delta_2, ...) \in T_i \triangleq \times_{k=0}^{\infty} \Delta(X_k)$ is called a hierarchy (of beliefs).

For the topology-free model we describe here, Heifetz and Samet (1998) prove the existence of a universal type space comprised of a subset of hierarchies, $T_i^* \subset T_i$, and a belief, $\pi_i^* : T_i^* \to \Delta\left(T_{-i}^* \times \Theta\right)$, for all i. The type space is universal in that any other type space can be uniquely translated into this universal type space. Specifically for any hierarchy $t \in T^*$, we write $\delta_k(t)$ for the k^{th} component of t and we write T_k^* for the set of k^{th} level beliefs for all types in T^* , $T_k^* \subseteq \Delta(X_{k-1})$. (Where no confusion results we drop the subscript of i for notational simplicity.) Type t's marginal beliefs about the state of Nature are

$$\widehat{\pi}_i^1[t_i](\theta) = \hat{\pi}_i[t_i](\{(t_{-i}, \theta) : t_{-i} \in T_{-i}\}).$$

More generally, for each $k = 2, 3...$, the translation implies the existence of $\widehat{\pi}_i^k : T_i \to T_k^*$ such that for $E \subseteq \left(T_{k-1}^*\right)^{I-1} \times \Theta$,

$$\widehat{\pi}_i^k[t_i](E) = \widehat{\pi}_i[t_i]\left(\left\{t_{-i} \in T_{-i} : \left(\widehat{\pi}_{-i}^{k-1}(t_{-i}), \theta\right) \in E\right\}\right).$$

Let $\widehat{\pi}_i^*(t_i) = \left(\widehat{\pi}_i^k(t_i)\right)_{k=1}^{\infty}$, and then $\widehat{\pi}_i^* : T_i \to T_i^*$ is the translation discussed above.

2.1 Examples of “Redundant” Types

A player's type captures everything about his beliefs and higher order beliefs about Θ. However, type spaces also contain types that cannot be distinguished on the basis of their beliefs and higher order beliefs about Θ. While Mertens and Zamir (1985) labelled these “redundant types”, they may nonetheless be strategically relevant. This issue, and its significance for the interpretation of

[4] Note that if X is Polish then this belief-generated field corresponds to the Borel field when X is endowed with the topology of weak convergence of measures (i.e. the "weak" topology) Thus we also use $\Delta(X)$ to denote the Borel field when it is equivalent.

[5] See also Armbruster and Boge (1979), Brandenburger and Dekel (1993), and Heifetz (1993), among others.

the universal type space, has been discussed by Bergemann and Morris (2001) and Battigalli and Siniscalchi (2003). In particular, it is not true that

$$\widehat{\pi}_i^*\left(t_i'\right)=\widehat{\pi}_i^*\left(t_i\right)\Rightarrow\widehat{\pi}_i\left(t_i'\right)=\widehat{\pi}_i\left(t_i\right).$$

As noted by the above authors, this is most easy to see in the case of complete information, where Θ is a singleton.

Example 1 *Let $I=2$, $T_1=\{t_1,t_1',t_1''\}$, $T_2=\{t_2,t_2',t_2''\}$ and $\Theta=\{\theta\}$. Let beliefs be generated by the common prior below. Call this type space $\mathcal{T}^1$.*

θ	t_2	t_2'	t_2''
t_1	$\frac{1}{6}$	$\frac{1}{12}$	0
t_1'	$\frac{1}{12}$	$\frac{1}{6}$	0
t_1''	0	0	$\frac{1}{2}$

Now (for each i) $\widehat{\pi}_i^*\left(t_i''\right)=\widehat{\pi}_i^*\left(t_i'\right)=\widehat{\pi}_i^*\left(t_i\right)$ but $\widehat{\pi}_i\left(t_i'\right)\neq\widehat{\pi}_i\left(t_i\right)$, $\widehat{\pi}_i\left(t_i''\right)\neq\widehat{\pi}_i\left(t_i\right)$ and $\widehat{\pi}_i\left(t_i''\right)\neq\widehat{\pi}_i\left(t_i'\right)$.

We will see that in defining rationalizability, this particular type of redundancy is relatively easy to deal with. But the redundancy in the following type space turns out to be trickier.

Example 2 *Let $I=2$, $T_1=\{t_1,t_1',t_1''\}$, $T_2=\{t_2,t_2',t_2''\}$ and $\Theta=\left\{\theta,\theta'\right\}$. Let beliefs be generated by a common prior below. Call this type space $\mathcal{T}^2$.*

θ	t_2	t_2'	t_2''
t_1	$\frac{1}{12}$	$\frac{1}{24}$	0
t_1'	$\frac{1}{24}$	$\frac{1}{12}$	0
t_1''	0	0	$\frac{1}{4}$

θ'	t_2	t_2'	t_2''
t_1	$\frac{1}{24}$	$\frac{1}{12}$	0
t_1'	$\frac{1}{12}$	$\frac{1}{24}$	0
t_1''	0	0	$\frac{1}{4}$

Again (for each i) $\widehat{\pi}_i^*\left(t_i''\right)=\widehat{\pi}_i^*\left(t_i'\right)=\widehat{\pi}_i^*\left(t_i\right)$ but $\widehat{\pi}_i\left(t_i'\right)\neq\widehat{\pi}_i\left(t_i\right)$, $\widehat{\pi}_i\left(t_i''\right)\neq\widehat{\pi}_i\left(t_i\right)$ and $\widehat{\pi}_i\left(t_i''\right)\neq\widehat{\pi}_i\left(t_i'\right)$.

3 Games and solution concepts

Each player has a measure space of possible actions A_i. A game g consists of, for each player, a payoff function g_i, where $g_i: A\times\Theta\rightarrow[0,1]$. Write $\mathcal{G}$ for the set of possible games. The solution concepts we study are applied to a pair $(g,\mathcal{T})$, and specify possible action profiles for such a game of incomplete information, where an action profile is denoted by $a=(a_i)_{i\in I}\in\Pi_i\left(A^{T_i}\right)$.

Our main solution concept is ε (correlated) interim rationalizability, where ε is a measure of sub-optimization. To clarify the role of correlation we also provide a definition of independent interim rationalizability. We also define other solution concepts and equivalencies among them in a manner that is analogous to what is known for the case of complete information (i.e., Θ being a singleton).

We begin with a fixed-point definition of the solution concepts, because in the general environment we allow the usual iterative process may require transfinite induction; see Lipman (1994). We then prove that when A and Θ are finite this fixed point corresponds to definition using a standard iteration procedure. The section ends with examples studying how these concepts may depend on (more than) just the hierarchy of beliefs in a type space. Throughout we hold the game g and the number ε fixed; hence to simplify notation and terminology we do not explicitly write that various functions depend on these parameters, e.g. the phrase "best reply" will mean a reply that gives within ε of the maximum payoff. The main question is whether the solution depends on the type space, so we do specify the dependence on $\mathcal{T}$.

3.1 Best replies and undominated actions

For any subset of actions for all types , we first define the interim rationalizable actions when beliefs over opponents are restricted to those actions.

Definition 1 *The correspondence of best replies for all types given a subset of actions for all types is denoted* $\mathrm{BR}^{\mathcal{T}} : \left(\left(2^{A_i}\right)^{T_i}\right)_{i\in I} \rightarrow \left(\left(2^{A_i}\right)^{T_i}\right)_{i\in I}$ *and is defined as follows. First, given* $E_{-i} = \left(\left(E_{t_j}\right)_{t_j\in T_j}\right)_{j\neq i}$, *with non-empty* $E_{t_j} \subset A_{t_j}$ *for all* t_j *and* $j \neq i$, *we define the* ε *best replies for* t_i *in game* g *as*

$$BR_i^{\mathcal{T}}\left(t_i, E_{-i}\right) = \left\{ a_i \in A_i \;\middle|\; \begin{array}{l} \exists \nu \in \Delta\left(T_{-i} \times \Theta \times A_j\right) \textit{ such that} \\ (\mathrm{i}) \; \nu\left[\left\{\left(t_{-i}, \theta, a_{-i}\right) : a_{t_j} \in E_{t_j} \textit{ for all } j \neq i\right\}\right] = 1 \\ (\mathrm{ii}) \; marg_{T_{-i}\times\Theta}\nu = \widehat{\pi}_i\left(t_i\right) \\ (\mathrm{iii}) \; \int\limits_{\left(t_{-i},\theta,a_{-i}\right)} \left[\begin{array}{c} g_i\left(a_i, a_{-i}, \theta\right) \\ -g_i\left(a_i', a_{-i}, \theta\right) \end{array}\right] d\nu \geq -\varepsilon \textit{ for all } a_i' \in A_i \end{array} \right\} \tag{1}$$

Next, given $E = \left(\left(E_{t_i}\right)_{t_i\in T_i}\right)_{i\in I} \subset \left(A^{T_i}\right)_{i\in I}$, *we define* $BR^{\mathcal{T}}\left(E\right) = \left(\left(BR_i^{\mathcal{T}}\left(t_i, E_{-i}\right)\right)_{t_i\in T_i}\right)_{i\in I}$. [6]

To define the independen*t* interim best replies we append to the conditions defining BR the following requirement: *(iv)* $\forall j \neq i, \exists \alpha_j : T \rightarrow \Delta\left(A_j\right)$ s.t. $\nu\left(F \times \left\{\left(a_j\right)_{j\neq i}\right\}\right) = \int_F \Pi_{j\neq i}\alpha_j\left(a_j\right) \mathrm{d}\hat{\pi}_i\left(t_{-i}, \theta\right)$.

[6] We absue notation and write BR both for the corresondence specifying best replies for a type and for the corresondence specifying these actions for all types.

We denote this function and the resulting fixed point by adding the prefix I, thus the best reply correspondence is denoted $\mathrm{IBR}^{\mathcal{T}}$.

Condition (iv) implies condition (ii), by adding up over all a_{-i}, but we state it separately to facilitate providing (and comparing with) the main definition that follows. Note that (iv) embodies two forms of independence: beliefs over opponents' actions are determined by multiplying them, and opponents' actions are independent of Θ conditional on their type.

Fudenberg and Tirole (1991, page 226) demonstrate the important distinction between interim and ex ante (strictly) dominated strategies. By the standard duality argument showing the equivalence of strict domination and never best response, we can define a correspondence of undominated strategies that is equivalent to $BR^{\mathcal{T}}$.

Specifically, given $E_{-i} = \left(\left(E_{t_j}\right)_{t_j \in T_j}\right)_{j \neq i}$, with non-empty $E_{t_j} \subset A_{t_j}$ for all t_j and $j \neq i$, the ε interim undominated actions for t_i in game g are $U_i^{\mathcal{T}}\left(t_i, E_{-i}\right)$ defined below.

$$U_i^{\mathcal{T}}\left(t_i, E_{-i}\right) = \left\{ a_i \in A_i \left| \begin{array}{l} \text{There does not exist } \alpha_i \in \Delta\left(A_i\right) \text{ such that,} \\ \text{for all } s_{-i} \in E_{-i}, \\ \int\limits_{\left(t_{-i},\theta\right)} \left[\begin{array}{c} \sum\limits_{a_i'} \alpha_i\left(a_i'\right) g_i\left(a_i', s_{-i}\left(t_{-i}\right), \theta\right) \\ -g_i\left(a_i, s_{-i}\left(t_{-i}\right), \theta\right) \end{array} \right] \widehat{d\pi_i}\left[t_i\right] > \varepsilon \end{array} \right. \right\}$$

The resulting correspondence, $U_i^{\mathcal{T}} : \left(\left(2^{A_i}\right)^{T_i}\right)_{i \in I} \to \left(\left(2^{A_i}\right)^{T_i}\right)_{i \in I}$ is defined analogously to $BR^{\mathcal{T}}$, and the equivalence $U_i^{\mathcal{T}} = BR_i^{\mathcal{T}}$ follows from standard arguments.

3.2 Fixed-point definitions

Definition 2 *The set of interim rationalizable actions (for all types) is the largest fixed point of* $\mathrm{BR}^{\mathcal{T}}$; *we denote this set by* $R^{\mathcal{T}} = \left(\left(R_i^{\mathcal{T}}\left(t_i\right)\right)_{t_i \in T_i}\right)_{i \in I} \subset \left(A^{T_i}\right)_{i \in I}$ *(* $\mathrm{BR}^{\mathcal{T}}$ *is a decreasing function on a complete lattice, so the largest fixed point exists)*

We can similarly define the equivalent fixed point of undominated actions, $\mathcal{U}^{\mathcal{T}} = R^{\mathcal{T}}$. The profile of independent interim rationalizable actions is analogously the largest fixed point of the decreasing function *IBR*, and is denoted $IR^{\mathcal{T}} = \left(\left(IR_i^{\mathcal{T}}\left(t_i\right)\right)_{t_i \in T_i}\right)_{i \in I}$.

We will want to use another alternative characterization of the interim rationalizable actions. Let $S_i^{\mathcal{T}} : T_i \to 2^{A_i} / \varnothing$ and $S^{\mathcal{T}} = \left(S_1^{\mathcal{T}}, ..., S_I^{\mathcal{T}}\right)$.

Definition 3 $S^{\mathcal{T}}$ *is a best-reply set if for each* t_i *and* $a_i \in S_i^{\mathcal{T}}\left(t_i\right)$, *there exists* $\nu \in \Delta\left(T_{-i} \times \Theta \times A_{-i}\right)$

such that

$$\begin{array}{l} \textit{(i)}\ \nu\left[\left\{(t_{-i},\theta,a_{-i}) : a_j \in S_j^{\mathcal{T}}(t_j)\ \textit{ for each } j \neq i\right\}\right] = 1 \\ \textit{(ii)}\ marg_{T_{-i}\times\Theta}\nu = \widehat{\pi}_i(t_i) \\ \textit{(iii)} \int\limits_{(t_{-i},\theta,a_{-i})} \left[\begin{array}{c} g_i(a_i,a_j,\theta) \\ -g_i(a_i',a_j,\theta) \end{array}\right] d\nu \geq -\varepsilon \textit{ for all } a_i' \in A_i \end{array}$$

The following two properties are now immediate from the definitions.

Lemma 1 *$R^{\mathcal{T}}$ is a best reply set.*

Lemma 2 *If $S^{\mathcal{T}}$ is a best reply set, then $S_i^{\mathcal{T}}(t_i) \subseteq R_i^{\mathcal{T}}(t_i)$ for all i and t_i.*

3.3 Iterative definitions

If the action and nature spaces, $A \times \Theta$, are finite, then instead of using a fixed point argument we could apply $BR^{\mathcal{T}}$ (or $U^{\mathcal{T}}$) iteratively. That is, let $R_1^{\mathcal{T}} = BR^{\mathcal{T}}(A)$, and $R_k^{\mathcal{T}} = BR^{\mathcal{T}}\left(R_{k-1}^{\mathcal{T}}\right)$, $R_\infty^{\mathcal{T}} = \cap_{k=1}^{\infty} R_k^{\mathcal{T}}$. Under such finiteness assumptions $R^{\mathcal{T}} = R_\infty^{\mathcal{T}}$. That this is true if $\mathcal{T}$ is finite is immediate.[7] But we now argue why it holds true of arbitrary $\mathcal{T}$. The key to the argument is that, regardless of the nature of the type space, the set $\Delta(A_{-i} \times \Theta)$ is a finite-dimensional metric space, and it is elements of this set that determine payoffs and best replies.

Proposition 1 *If $A \times \Theta$ is finite, then $R^{\mathcal{T}} = R_\infty^{\mathcal{T}}$.*

PROOF: We claim that $R_\infty^{\mathcal{T}}$ is a best-reply set, and therefore that the iterative process and fixed-point definition coincide. That nothing larger can be a best-reply set is immediate; the only question is whether after the taking the limit of the iterative procedure, further actions could be deleted. Put differently, we claim that $BR^{\mathcal{T}}\left(R_\infty^{\mathcal{T}}\right) = R_\infty^{\mathcal{T}}$.

Let $\Psi_i^k \subseteq \Delta(A_{-i} \times \Theta)$ be the set of beliefs over $A_{-i} \times \Theta$ consistent with $R_k^{\mathcal{T}}$, i.e., let

$$\Psi_i^k = \left\{ \psi \in \Delta(A_{-i} \times \Theta) : \begin{array}{l} \psi = \int_{T_{-i}} \nu(t_{-i},\theta,a_{-i})\,\mathrm{d}t_{-i} \text{ for some } \nu \in \Delta(T_{-i} \times \Theta \times A_j) \text{ s.t.} \\ (i)\ \nu\left[\left\{(t_{-i},\theta,a_{-i}) : a_{t_j} \in R_{j,k}^{\mathcal{T}}(t_j) \text{ for all } j \neq i\right\}\right] = 1 \\ (ii)\ \mathrm{marg}_{T_{-i}\times\Theta}\nu = \widehat{\pi}_i(t_i) \\ (iii) \int\limits_{(t_{-i},\theta,a_{-i})} \left[\begin{array}{c} g_i(a_i,a_{-i},\theta) \\ -g_i(a_i',a_{-i},\theta) \end{array}\right] \mathrm{d}\nu \geq -\varepsilon \text{ for all } a_i' \in A_i \end{array} \right\}.$$

[7] We conjecture it also holds for type spaces (T,π) where T , Θ and A are topological spaces, and where g is continuous and π is continuous and maps into regular measures on $\Delta(\Theta \times T_{-i})$.

If a_i can be deleted by transfinite induction for type t_i, that is, if $BR^{\mathcal{T}}\left(R_\infty^{\mathcal{T}}\right) \subsetneq R_\infty^{\mathcal{T}}$, then there does not exist $\psi \in \Psi_i^\infty$ such that a_i is a best reply for t_i against ψ. That is, for each $\psi \in \Psi_i^\infty$ there exists $a_i' \in A_i$ s.t. $\int\limits_{(\theta,a_{-i})} [g_i(a_i, a_{-i}, \theta) - g_i(a_i', a_{-i}, \theta)]\mathrm{d}\psi < -\varepsilon$.

By the duality between strict dominance and never-best-replies, there exists $\alpha_i \in \Delta(A_i)$ such that $\int\limits_{(\theta,a_{-i})} [g_i(a_i, a_{-i}, \theta) - g_i(\alpha_i, a_{-i}, \theta)]\, d\psi < -\varepsilon$ for all $\psi \in \Psi_i^\infty$. By continuity of expected utility there is a neighborhood of ψ such that this remains true. And because $\Delta(A_{-i} \times \Theta)$ is a compact metric space, and the Ψ_i^k are a decreasing sequence of subsets converging to Ψ_i^∞, there is a finite k such that $\int\limits_{(\theta,a_{-i})} [g_i(a_i, a_{-i}, \theta) - g_i(\alpha_i, a_{-i}, \theta)]\, d\psi \leq -\varepsilon$ for all $\psi \in \Psi_i^k$. Hence $a_i \notin R_k^{\mathcal{T}}(t_i)$, leading to a contradiction. □

3.4 Examples

In example 1, we clearly have $IR_i^{\mathcal{T}^1}(t_i) = IR_i^{\mathcal{T}^1}(t_i') = IR_i^{\mathcal{T}^1}(t_i'')$ and $R_i^{\mathcal{T}^1}(t_i) = R_i^{\mathcal{T}^1}(t_i') = R_i^{\mathcal{T}^1}(t_i'')$ for any two-player game g. In particular, these sets will be the ε-rationalizable actions of the underlying complete-information game and the result is an implication of the equivalence of correlated and independent rationalizability in two-player complete-information games.

But in the type space of example 2, things wont be so simple, even in two-player games. Consider the following game g, where player 1 chooses the row and player 2 chooses the column.

θ	L	R
u	$1, 0$	$0, 0$
d	$\frac{3}{4}, 0$	$\frac{3}{4}, 0$

θ'	L	R
u	$0, 0$	$1, 0$
d	$\frac{3}{4}, 0$	$\frac{3}{4}, 0$

Let $\varepsilon = 0$. Now $IR_2^{\mathcal{T}^2}(t_2) = IR_2^{\mathcal{T}^2}(t_2') = IR_2^{\mathcal{T}^2}(t_2'') = \{L, R\}$; and $IR_1^{\mathcal{T}^2}(t_1) = IR_1^{\mathcal{T}^2}(t_1') = \{u, d\}$, but $IR_1^{\mathcal{T}^2}(t_1'') = \{d\}$. However, $R_2^{\mathcal{T}^2}(t_2) = R_2^{\mathcal{T}^2}(t_2') = R_2^{\mathcal{T}^2}(t_2'') = \{L, R\}$; and $R_1^{\mathcal{T}^2}(t_1) = R_1^{\mathcal{T}^2}(t_1') = R_1^{\mathcal{T}^2}(t_1'') = \{u, d\}$. To see why $u \in R_1^{\mathcal{T}^2}(t_1'')$, let type t_1'' put probability $\frac{1}{2}$ on (t_2'', θ, L) and probability $\frac{1}{2}$ on $\left(t_2'', \theta', R\right)$.

This example has the same flavor as examples showing the non-equivalence of correlated and interim rationalizability in three-player complete-information games. For example, consider the three-player game where player 1 chooses the row, player 2 chooses the column and player 3 chooses the matrix, with payoffs

A	L	R
u	$1, 0, 0$	$0, 0, 0$
d	$\frac{3}{4}, 0, 0$	$\frac{3}{4}, 0, 0$

B	L	R
u	$0, 0, 0$	$1, 0, 0$
d	$\frac{3}{4}, 0, 0$	$\frac{3}{4}, 0, 0$

Here, action d fails to be independently rationalizable for player 1 but is correlated rationalizable. In an influential argument, Aumann (1987) writes in this context that

> ...in games with more than two players, correlation may express the fact that what 3, say, thinks that 1 will do may depend on what he thinks 2 will do. This is no connection with any overt or even covert collusion between 1 and 2; they may be acting entirely independently....(page 612)

We propose treating nature as another player. If player 1, say, does not know what determines which of his rationalizable actions player 2 will play, why should this subjective uncertainty be completely independent of the uncertainty about the choice of nature? This interpretation introduces the possibility that there are other (payoff irrelevant) states of the world that are not modelled in Θ but that lead to these beliefs. We explicitly exploit such an expansion of the space in Section 5 to prove the equivalence of interim rationalizability with more familiar solution concepts.

4 Measurability of Interim Rationalizable Sets

In the last example, the set of independent interim rationalizable actions depended not only on a type's beliefs and higher order beliefs about Θ, but also on the type space within which that type was embedded. But the set of interim rationalizable actions depended only on a type's beliefs and higher-order beliefs about Θ. The following proposition shows that this is true in general.

Proposition 2 *Given two type spaces on the basic set of states Θ, $\mathcal{T}$ and $\mathcal{T}$, with t_i a type of i in $\mathcal{T}$ and t_i' a type of i in $\mathcal{T}'$, we have* $\widehat{\pi}_i^*\left(t_i'\right) = \widehat{\pi}_i^*\left(t_i\right) \Rightarrow R_i^{\mathcal{T}'}\left(t_i'\right) = R_i^{\mathcal{T}}\left(t_i\right).$

.

PROOF: (i)We will prove this using the ε-best-reply sets. Specifically, consider a best-reply set $S^{\mathcal{T}}$, and recall the translation of $\mathcal{T}$ into hierarchies of beliefs discussed in section 2, $\hat{\pi}^* : T \to T^*$. The claim is that if we replace all types by their image in the universal type space we still obtain a best-reply set. That is, map t_i into $t_i^* = \hat{\pi}_i^*\left(t_i\right)$, and $S^{\mathcal{T}^*}\left(t_i^*\right) = \left\{a_i : a_i \in S^{\mathcal{T}}\left(t_i\right) \text{ for } t_i \in \left(\hat{\pi}_i^*\right)^{-1}\left(t_i^*\right)\right\}$. For any $\nu \in \Delta\left(T_{-i} \times \Theta \times A_{-i}\right)$ define $\nu^* \in \Delta\left(T_{-i}^* \times \Theta \times A_{-i}\right)$ by

$$\nu^*\left(E \times \{\theta\} \times \{a\}\right) = \nu\left(\left(\hat{\pi}^*\right)^{-1}(E) \times \{\theta\} \times \{a\}\right)$$

for $E \subset T_{-i}^*$. It is immediate that conditions (i)–(iii) in the definition of best-reply sets are satisfied. Thus we have shown that for any type space $\mathcal{T}$ and any type t_i of i in that space,

$S(t_i) \subset S(\hat{\pi}^*(t_i))$. The converse is obtained similarly. For any $\nu^* \in \Delta\left(T^*_{-i} \times \Theta \times A_{-i}\right)$ define $\nu \in \Delta(T_{-i} \times \Theta \times A_{-i})$ by

$$\nu(E \times \{\theta\} \times \{a\}) = \nu^*(\hat{\pi}^*(E) \times \{\theta\} \times \{a\})$$

for $E \subset T_{-i}$. Again (i)–(iii) are immediate. Therefore $S(t_i) = S(\hat{\pi}^*(t_i))$. $\square$

Claim: One can also show that $\widehat{\pi}_i^k(t_i') = \widehat{\pi}_i^k(t_i) \Rightarrow R_{i,k}^{\mathcal{T}'}(t_i') = R_{i,k}^{\mathcal{T}}(t_i)$ for any finite k. We will provide a proof in a future version of this paper.

5 Interim Rationalizability, Equilibrium on Large Type Spaces and Common Knowledge of Rationality

One message from Brandenburger and Dekel (1987) was that equilibrium has no bite when there are large type spaces and the common prior assumption is dropped. We can state the analogous result for this incomplete information setting. Specifically, we prove that given any type space and game, any interim rationalizable action is also played in an equilibrium of that same game but with an expanded type space. Brandenburger and Dekel prove that any rationalizable action of a complete information game is played in some subjective correlated equilibrium, which is just an equilibrium of a game with an expanded type space that functions as a subjective correlating device. Our construction below is very similar, we expand the type spaces by adding to each player's type a signal that corresponds to a recommended action.

Fix type space $\mathcal{T}$. We will consider an enlarged type space ($\widetilde{\mathcal{T}} = \left(\widetilde{T}_i, \tilde{\pi}_i\right)_{i=1}^{I}$) which can be translated into $\mathcal{T}$ (with translation $\varphi_i : \widetilde{T}_i \rightarrow T_i$). Given a g and the type space $\widetilde{\mathcal{T}}$, we have an incomplete information game. A strategy profile $s = (s_1, ..., s_I)$, each $s_i : \widetilde{T}_i \rightarrow A_i$, measurable, is a pure strategy ε-interim equilibrium of the game $\left(g, \widetilde{\mathcal{T}}\right)$ if and only if

$$\begin{aligned} & \int\limits_{\widetilde{t}_{-i},\theta} g_i\left(s_i\left(\widetilde{t}_i\right), s_{-i}\left(\widetilde{t}_{-i}\right), \theta\right) d\widetilde{\pi}_i\left(\cdot|\widetilde{t}_i\right) \\ \geq & \int\limits_{\widetilde{t}_{-i},\theta} g_i\left(a_i, s_{-i}\left(\widetilde{t}_{-i}\right), \theta\right) d\widetilde{\pi}_i\left(\cdot|\widetilde{t}_i\right) - \varepsilon \end{aligned}$$

for all i, $\widetilde{t}_i \in \widetilde{T}_i$ and $a_i \in A_i$.

Proposition 3 *$\bar{a}_i \in R_i^{\mathcal{T}}(\bar{t}_i)$ if and only if there exists an enlarged type space $\widetilde{\mathcal{T}}$ and an ε-interim equilibrium of the game $\left(g, \widetilde{\mathcal{T}}\right)$, such that $s_i\left(\widetilde{t}_i\right) = \bar{a}_i$ and $\varphi_i\left(\widetilde{t}_i\right) = \bar{t}_i$ for some $\widetilde{t}_i \in \widetilde{T}_i$.*

PROOF. For each $a_i \in R^{\mathcal{T}}(t_i)$, by Lemma 1, there exists $\nu_{a_i,t_i} \in \Delta(T_{-i} \times \Theta \times A_{-i})$ such that

$$\nu_{a_i,t_i}\left[\left\{(t_{-i},\theta,a_{-i}) : a_j \in R^{\mathcal{T}}(t_j) \text{ for all } j \neq i\right\}\right] = 1,$$

$$\text{marg}_{T_{-i}\times\Theta}\nu_{a_i,t_i} = \widehat{\pi}_i(t_i)$$

and

$$\int\limits_{(t_{-i},\theta,a_{-i})} \left[\begin{array}{c} g_i(a_i,a_{-i},\theta) \\ -g_i(a_i',a_{-i},\theta) \end{array}\right] d\nu_{a_i,t_i} \geq -\varepsilon \text{ for all } a_i' \in A_i. \tag{2}$$

Now consider the following enlarged type space with

$$\begin{array}{rcl} \tilde{T}_i &=& T_i \times A_i \\ \widetilde{\pi}_i(\cdot|(t_i,a_i)) &=& \nu_{a_i,t_i}, \\ \varphi_i((t_i,a_i)) &=& t_i. \end{array}$$

Consider the strategy profile with

$$s_i((t_i,a_i)) = a_i.$$

By construction, if $\widetilde{t}_i = (t_i,a_i)$,

$$\begin{array}{rl} & \int\limits_{\widetilde{t}_{-i},\theta} g_i\left(s_i\left(\widetilde{t}_i\right), s_{-i}\left(\widetilde{t}_{-i}\right),\theta\right) d\widetilde{\pi}_i\left(\cdot|\widetilde{t}_i\right) \\ = & \int\limits_{\widetilde{t}_{-i},\theta} g_i\left(s_i\left(\widetilde{t}_i\right), s_{-i}\left(\widetilde{t}_{-i}\right),\theta\right) d\nu_{a_i,t_i} \\ \geq & \int\limits_{\widetilde{t}_{-i},\theta} g_i\left(a_i', s_{-i}\left(\widetilde{t}_{-i}\right),\theta\right) d\nu_{a_i,t_i} - \varepsilon \\ = & \int\limits_{\widetilde{t}_{-i},\theta} g_i\left(a_i', s_{-i}\left(\widetilde{t}_{-i}\right),\theta\right) d\widetilde{\pi}_i\left(\cdot|\widetilde{t}_i\right) - \varepsilon \end{array}$$

for all a_i'.

Conversely, suppose that there exists an enlarged type space $\widetilde{\mathcal{T}}$ and an ε-interim equilibrium of the game $\left(g,\widetilde{\mathcal{T}}\right)$, s. Let

$$S_i(t_i) = \left\{a_i : s_i\left(\widetilde{t}_i\right) = a_i \text{ and } \varphi_i\left(\widetilde{t}_i\right) = t_i \text{ for some } \widetilde{t}_i \in \mathcal{T}_i\right\}.$$

Suppose $s_i\left(\widetilde{t}_i\right) = \bar{a}_i$ and $\varphi_i\left(\widetilde{t}_i\right) = \bar{t}_i$ for some $\widetilde{t}_i \in \mathcal{T}_i$. Since S satisfies the $\varepsilon-$best response property for game g, we have by Lemma 2 that $\bar{a}_i \in S_i(\bar{t}_i) \subseteq R_i^{\mathcal{T}}(\bar{t}_i)$. □

Remark 1 *Following Bernheim (1984) and Pearce (1984) we can reinterpret proposition 2 as establishing that the set of interim rationalizable actions for type t_i are the set of actions that are consistent with common knowledge of rationality (while maintaining the implicit assumption that there is common knowledge of the game g and type t_i's beliefs and higher order beliefs about Θ). Similarly we can follow Aumann (1987), Brandenburger and Dekel (1987) and Tan and Werlang (1988) and view proposition 3 as relating common knowledge of rationality to a solution concept. Specifically, suppose we interpret the function $s_i : \tilde{T}_i \to A_i$ on the enlarged state space as representing the action which is played as a function of a player's type. If we fix the game g, we can define the event that player i is rational given payoff function g_i.*

$$[Rat_i\,(g,\varepsilon)] = \left\{ \tilde{t}_i : s_i\left(\tilde{t}_i\right) \in \left\{ a_i \left| \int\limits_{\left(\tilde{t}_{-i},\theta\right)} \begin{bmatrix} g_i\left(a_i, s_{-i}\left(\widetilde{t}_{-i}\right), \theta\right) \\ -g_i\left(a_i', s_{-i}\left(\widetilde{t}_{-i}\right), \theta\right) \end{bmatrix} d\widetilde{\pi}_i\left(\widetilde{t}_{-i}, \theta | \widetilde{t}_i\right) \geq -\varepsilon,\ \forall a_i' \right. \right\} \right\}$$

Common knowledge of rationality holds only on a belief-closed space contained in $[Rat_i\,(g,\varepsilon)]$ for all i. Thus Proposition 3 establishes that common knowledge of rationality implies that players choose interim rationalizable actions and that any interim rationalizable action is consistent with common knowledge of rationality.

6 References

References

[1] Armbruster, A. and W. Böge (1979). "Bayesian Game Theory," in *Game Theory and Related Topics,* edited by O. Moeschlin and D. Pallaschke (Amsterdam: North-Holland), 17–28.

[2] Aumann, R. (1987). "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica*, **55**, 1-18.

[3] Battigalli, P. and M. Siniscalchi (2003). "Rationalization and Incomplete Information," *Advances in Theoretical Economics* **3** (1) Article 3. http://www.bepress.com/bejte/advances/vol3/iss1/art3/.

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

[5] Bergemann, D. and S. Morris (2001). "Robust Mechanism Design," http://www.econ.yale.edu/~sm326/rmd-nov2001.pdf.

[6] Brandenburger, A. and E. Dekel (1987). “Rationalizability and Correlated Equilibria,” *Econometrica*, **55**, 1391-1402.

[7] Brandenburger, A. and E. Dekel (1993). "Hierarchies of Beliefs and Common Knowledge," *Journal of Economic Theory* **59**, 189-198.

[8] Dekel, E., D. Fudenberg and S. Morris (2004). "Topologies on Types."

[9] Ely, J. and M. Peski (2004). "Hierarchies of Belief and Interim Rationalizability," http://papers.ssrn.com/sol3/papers.cfm?abstract_id=626641.

[10] Forges, F. (1993). "Five Legitimate Definitions of Correlated Equilibrium in Games with Incomplete Information," *Theory and Decision* **35**, 277-310.

[11] Fudenberg, D. and J. Tirole (1991). *Game Theory* Cambridge MA: MIT Press.

[12] Harsanyi, J. C. (1967-8). "Games with Incomplete Information Played by ‘Bayesian’ Players," parts I, II, and III, Management Science **14**, 159-182, 320-334, and 486-502.

[13] Heifetz, A. (1993). "The Bayesian Formulation of Incomplete Information—The Non-Compact Case," *The International Journal of Game Theory* **21**, 329-338.

[14] Heifetz A., and D. Samet (1998), "Topology-Free Typology of Beliefs" *Journal of Economic Theory* **82**, 324-341.

[15] Lipman, B. (1994), "A Note on the Implications of Common Knowledge of Rationality," *Games and Economic Behavior* **6**, 114-129.

[16] Mertens, J.-F., S. Sorin and S. Zamir (1994). "Repeated Games: Part A Background Material," CORE Discussion Paper #9420.

[17] Mertens, J.-F. and S. Zamir (1985). "Formulation of Bayesian Analysis for Games with Incomplete Information," *International Journal of Game Theory* **14**, 1-29.

[18] Pearce, D. "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica* **52**, 1029-1051.

[19] Tan, T. and S. Werlang (1988), “The Bayesian Foundation of Solution Concepts of Games,” *Journal of Economic Theory* **45**, 370-391.

[20] Weinstein, J. and M. Yildiz (2003). “Finite Order Implications of Any Equilibrium,” http://econ-www.mit.edu/faculty/download_pdf.php?id=911.