Interim Correlated Rationalizability[1]

Eddie Dekel
Northwestern University
and Tel Aviv University

Drew Fudenberg
Harvard University

Stephen Morris
Princeton University

First Draft: May 2003. This Draft: February 2006

---

[1] This paper began as part of our work on "Topologies on Types" but took on a life of its own in 2003. We thank seminar participants in many audiences since then for helpful comments, and are particularly grateful to Aviad Heifetz and Muhamet Yildiz for helpful discussions, Marcin Peski for detailed comments and Satoru Takahashi for expert proofreading. We acknowledge financial support from the National Science Foundation under grants 0111180, 0426199 and 0518929. Morris is grateful for financial support from the John Simon Guggenheim Foundation and the Center for Advanced Studies in the Behavioral Sciences.

Abstract

This paper proposes the solution concept of interim correlated rationalizability, and shows that all types that have the same hierarchies of beliefs have the same set of interim correlated 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. However, 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 the type spaces are mapped into the same subset of the universal type space.[1] In response, this paper proposes the solution concept of interim correlated rationalizability.

We show that the concept is well-defined, that its iterative and fixed-point definitions coincide, and that any two types with the same hierarchy of beliefs have the same interim-correlated-rationalizable actions, regardless of whether they reside in the same type space. Thus this is a solution concept that can be characterized by working with the universal type space.[2] We also show that the solution concept has similar properties to its complete-information counterpart. First, in claim 1, we note that the process of iterative elimination of interim strictly dominated strategies yields the same solution. Second, proposition 2 shows that interim correlated rationalizability is characterized by common knowledge of rationality. Third, we extend a result of Brandenburger and Dekel (1987). They 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; remark 2 reports a straightforward extension of Brandenburger and Dekel's observation to games with incomplete information.

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. Our focus is on the concept of interim correlated rationalizability, but we also define the concept of interim independent rationalizability; we use comparisons between the two concepts to help explain

---

[1] See Bergemann and Morris (2005) and Battigalli and Siniscalchi (2003) for a discussion of this issue.

[2] We use the concept of interim correlated rationalizability in our study of topologies on the universal type space (Dekel, Fudenberg and Morris (2006)). For that purpose, it is important to know that the solution concept depends only on hierarchies of beliefs (and not on other, ”redundant,” elements of the type space), as we establish here.

and motivate the correlated version. These two solution concepts are incomplete-information analogs of the complete-information concepts of correlated rationalizability and independent rationalizability, and reduce to them when Θ is a singleton. To understand these concepts, recall that all rationalizability notions involve iteratively deleting every action that is not a best reply to some player's beliefs, where at each stage of the deletion the beliefs are restricted to assign positive probability only to actions that have not yet been deleted. Our definitions of interim rationalizability iteratively delete actions for all types that are not best replies to some joint distribution on actions and states that is consistent with the beliefs of each type of each player about Θ and the other players' types, and with the restrictions on conjectures about the opponents' actions that were obtained at earlier stages of the iteration. In the case of independent interim rationalizability, the joint distributions considered possible by a player of type t are given by combining (independent) conjectures of strategy profiles for each opponent's types, with that type t's beliefs over opponents' types and over Θ. In the case of correlated interim rationalizability, t's beliefs over types and Θ are combined with any, perhaps correlated, conditional conjectures about which (surviving) actions are played at a given type profile and payoff relevant state. In this latter definition, a type's beliefs allow for correlation among the payoff-relevant state and other players' actions.[3]

Much work in this paper is devoted to establishing that our results hold on general type spaces. This generality is important for evaluating the claim that all types that map to the same point in the universal type space have the same set of interim-correlated-rationalizable outcomes, so that interim correlated rationalizability can be analyzed using the universal type space.

However, working on general type spaces introduces a number of technical complications, starting with the question of what sorts of type spaces to consider (we use the non-topological definition of Heifetz and Samet (1998)) and proceeding to the question of whether the set of best responses is measurable, whether transfinite induction is required to equate the iterative and fixed-point definitions of rationalizability, and so on. These issues are important for a general analysis, but they seem to shed little light on either the motivation for the definition

---

[3] 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, interim independent and correlated rationalizability 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.

of interim correlated rationalizability or its invariance property. For this reason we restrict attention to finite type spaces in the first part of the paper, which allows us to give less technical definitions and statements of some of our results. We then proceed to the more general analysis.

We now consider an example to illustrate some of these ideas. The example illustrates our conclusion that this concept corresponds to common knowledge of rationality and that it depends only on the types (hierarchies of beliefs) and not on other (redundant) aspects of the type space, and that the latter independence is not true for interim independent rationalizability. It also emphasizes the form of correlation allowed by our main concept; a more detailed discussion of this correlation appears in subsection 3.2.

Example 1 (The Effect of Correlation with Nature) *Consider the following two-player game with incomplete information,* Γ*. Player* 1 *chooses the row, player* 2 *chooses the column and Nature chooses whether payoffs are given by the left hand matrix (in state* θ*) or the right hand matrix (in state* θ'*).*

θ	L	R	θ'	L	R
U	$1, 0$	$0, 0$	U	$0, 0$	$1, 0$
D	$\frac{3}{5}, 0$	$\frac{3}{5}, 0$	D	$\frac{3}{5}, 0$	$\frac{3}{5}, 0$

We assume that each player believes that each state is equally likely, and that this is common knowledge.[4] *Clearly, either action is rational for player* 2*, as she is indifferent between all actions. Now suppose that player* 1 *believes that with probability* $\frac{1}{2}$*, the true state will be* θ *and player* 2 *will choose* L*, and with probability* $\frac{1}{2}$*, the true state will be* θ' *and player* 2 *will choose* R*. This belief makes* U *optimal for player 1. As we will see in section 3.4, this means that* U *is consistent with common knowledge of rationality.*

Every action is also consistent with correlated interim rationalizability. To illustrate this we consider two type spaces that capture the same assumptions as above about players' higher-order beliefs. In type space $\mathcal{T}$*, each player* $i = 1, 2$ *has two possible types,* $T_i = \{t'_i, t''_i\}$ *and*

---

[4] Formally, this means that the event that each player assigns equal probability to the states is common knowledge, as defined in section 3.4.

beliefs are generated by the following common prior over $T_1 \times T_2 \times \{\theta, \theta'\}$*:*

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

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

In $\hat{\mathcal{T}}$ *each player has one possible type, and the beliefs are given by the following common prior.*

θ	$\hat{t}_2$
$\hat{t}_1$	$\frac{1}{2}$

θ'	$\hat{t}_2$
$\hat{t}_1$	$\frac{1}{2}$

Notice that in both type spaces, for every type of both players, there is common knowledge that each player assigns probability $1/2$ *to the true state being* θ*. The types in* $\mathcal{T}$ *are redundant in the sense of Mertens and Zamir (1985): there are multiple copies of types that agree with respect to their beliefs and higher order beliefs about* θ*. But these types nonetheless differ in their beliefs about their opponents and this is potentially important; types* t_i' *and* t_i'' *differ from type* $\hat{t}_i$ *because their beliefs allow for correlation between the action of player* j *and the state.*

To find the interim correlated rationalizable actions of Γ *with the above type spaces we iteratively eliminate actions for each type* t_i *of player* i *that are not best responses to some conjecture for the player over the triples* (t_j, θ, a_j) *that puts probability on type action pairs* (t_j, a_j) *that have not been deleted and that are consistent with type* t_i*'s beliefs over* (t_j, θ)*. In the example, no action will be eliminated for any type in either type space by the argument that we gave above.*

Now consider the alternative solution concept of interim independent rationalizability, where we add the additional requirement that at each round, for an action to survive, type t_i*'s conjecture over* (t_j, θ, a_j) *must treat the choice of player* j*'s action as independent, conditional on his type. With this solution concept, action* U *will not be interim independent rationalizable for type* $\hat{t}_1$*: there is no conditionally independent conjecture that supports play of action* U*. Thus* D *is the only interim independent rationalizable action for type* $\hat{t}_1$*. On the other hand, if type* t_1' *thinks that type* t_2' *will play action* L *and type* t_2'' *will play action* R*, then he will attach probability* $\frac{1}{3}$ *to each of action-state profiles* (L, θ) *and* (R, θ')*. This is enough to make action* D *a best response. Thus both* U *and* D *are interim independent rationalizable for types* t_1' *and* t_1''*.*

The example shows how "redundant types" in the type space have no impact on the interim correlated rationalizable actions for a given type, but do matter for interim independent rationalizability. This result is proved and discussed in greater generality in the paper. The example also helps see the intuition for why our solution concept depends only on the types, and not the details of the type space: The concept allows players to have correlated conjectures over other players' actions and the state, so the ability of "redundant types" to support such correlation is, truly, redundant. In this sense, the classical universal type space of Mertens and Zamir (1985) is the "right" type space for our correlated version of interim rationalizability, for which the only part of a player's type that matters is his beliefs and higher order beliefs about θ.

There are three papers that study closely related issues. 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 "Δ" is set equal to a complete description of the infinite hierarchies of beliefs. With this Δ, our proposition 2 corresponds to their proposition 4.3. They do not analyze this particular "Δ" and therefore do not address the issue of the distinction between correlated and independent interim rationalizability.[5]

Forges (1993) examines different ways of defining correlated equilibrium for games of incomplete information. Her "universal Bayesian approach" (in section 6) allows a player's own actions to depend on the payoff states θ even when the player cannot distinguish between the states; this is analogous to the correlation in conjectures that we use in defining our solution concept (we discuss this further in subsection 3.2). Thus our approach is the non-common prior analogue of Forges' universal Bayesian approach.

A recent paper by Ely and Peski (2005) also notes that the set of interim independent rationalizable outcomes in two-player games depend on more than just the standard universal type space. In response, they provide an extended notion of hierarchies of beliefs for

---

[5] Their analysis deals with restrictions on first-order beliefs only. Our result thus extends their approach to a particular restriction on the entire hierarchy of beliefs.

two-player games, and show that interim independent rationalizability in two-player games depends on types only via those extended hierarchies.

2 Setup and Solution Concepts

The primitives of our model are a finite set Θ of states of Nature, a finite set of I players, $i = 1$ to I, a finite set of actions A_i for each player and a payoff function g_i, where $g_i : A \times \Theta \to [0, 1]$, and $A = (A_i)_{i \in I}$.[6] (Henceforth we use this notation, i.e., $Q = (Q_c)_{c \in C}$ for the ordered collection of any set of C sets $\{Q_c : c \in C\}$. Also, we use the index $j \neq i$ for $\{j \in I : j \neq i\}$ and write Q_{-i} for $(Q_j)_{j \neq i}$. Elements of these are written as usual as $q_c \in Q_c$, $q \in Q$, and $q_{-i} \in Q_{-i}$.) For the first part of the paper, we restrict attention to a finite type space $\mathcal{T} = (T_i, \pi_i)_{i=1}^{I}$, where each T_i is a finite set, and each π_i maps T_i to the set $\Delta(T_{-i} \times \Theta)$ of probability measures on the finite set $T_{-i} \times \Theta$.[7] We later relax the assumption that $\mathcal{T}$ is finite (and we do not repeat the restriction until then); its role here is to simplify issues regarding measurability and the choice of a sigma field.

Our view of this type space is that it is an exogenously given part of the model. This could be because the type space corresponds to some actual information structure, (but not necessarily one that is a complete description of the world—just whatever the modeler views as the "pertinent" parts) or is the modeler's (partial) description of the players' perception of the environment (the players' views of the beliefs about beliefs...about Θ). We discuss this further in section 3.4. In our description of the type space, and in the beliefs allowed in the solution concept described next, we do not restrict to common priors.

The main solution concept that we study is that of interim correlated rationalizability, or "ICR".[8] At each round of an iterative deletion procedure, an action survives for a given type only if it is a best response to a conjecture over $T_{-i} \times \Theta \times A_{-i}$ that (1) puts positive probability only on type-action pairs of the opponents that have not yet been deleted; and

---

[6] We abuse notation by also writing I for the set of players.

[7] Throughout the paper, every finite set is given the obvious sigma field.

[8] Yildiz (2006) uses the main argument from Weinstein and Yildiz (2003) to show that types in the universal type space with a unique ICR action are open and dense in the product topology. Dekel, Fudenberg and Morris (2006) define a strategic topology to be one that generates continuity of ICR actions. In this topology, there are open sets of types with multiple ICR actions and open sets of types with unique ICR actions.

(2) is consistent with that type's beliefs about $T_{-i} \times \Theta$. Formally, we have $R^{\mathcal{T}}_{i,0}(t_i) = A_i$;

$$R^{\mathcal{T}}_{i,k+1}(t_i) = \left\{ a_i \in A_i \,\middle|\, \begin{array}{l} \text{there exists } \nu \in \Delta(T_{-i} \times \Theta \times A_{-i}) \text{ such that} \\ (1)\ \nu[(t_{-i}, \theta, a_{-i})] > 0 \Rightarrow a_{-i} \in \left(R^{\mathcal{T}}_{j,k}(t_j)\right)_{j \neq i} \\ (2)\ \sum\limits_{A_{-i}} \nu[(t_{-i}, \theta, a_{-i})] = \pi_i(t_i)[(t_{-i}, \theta)] \\ (3)\ a_i \in \arg\max\limits_{a_i'} \sum\limits_{T_{-i} \times \Theta \times A_{-i}} g_i\left((a_i', a_{-i}), \theta\right) \nu[(t_{-i}, \theta, a_{-i})] \end{array} \right\};$$

and

$$R^{\mathcal{T}}_i(t_i) = \lim_{k \to \infty} R^{\mathcal{T}}_{i,k}(t_i).$$

To better explain ICR we will compare it to a related solution concept, interim independent rationalizability, or IIR. The latter concept imposes the additional restriction that type t_i's conjecture supporting an action corresponds to independent conjectures about each other player's action. Let $IIR^{\mathcal{T}}_{i,0}(t_i) = A_i$;

$$IIR^{\mathcal{T}}_{i,k+1}(t_i) = \left\{ a_i \in A_i \,\middle|\, \begin{array}{ll} \multicolumn{2}{l}{\text{there exists } \nu \in \Delta(T_{-i} \times \Theta \times A_{-i}) \text{ such that}} \\ \multicolumn{2}{l}{(1)\ \nu[(t_{-i}, \theta, a_{-i})] > 0 \Rightarrow a_{-i} \in \left(IIR^{\mathcal{T}}_{j,k}(t_j)\right)_{j \neq i}} \\ (2) & \begin{array}{c} \text{for each } j \neq i \text{ there exists } \sigma_j : T_j \to \Delta(A_j) \text{ such that} \\ \nu[(t_{-i}, \theta, a_{-i})] = \pi_i(t_i)[(t_{-i}, \theta)] \prod\limits_{j \neq i} \sigma_j(t_j)[a_j] \end{array} \\ \multicolumn{2}{l}{(3)\ a_i \in \arg\max\limits_{a_i'} \sum\limits_{T_{-i} \times \Theta \times A_{-i}} \nu[(t_{-i}, \theta, a_{-i})] g_i\left((a_i', a_{-i}), \theta\right)} \end{array} \right\};$$

and

$$IIR^{\mathcal{T}}_i(t_i) = \lim_{k \to \infty} IIR^{\mathcal{T}}_{i,k}(t_i).$$

Thus ICR and IIR can be seen as polar cases with respect to the amount and kind of correlation that is allowed. An intermediate concept, that we mention below but do not define formally, could allow for correlation among players but not with nature, by specifying $\sigma_{-i} : T_{-i} \to \Delta(A_{-i})$ instead of $(\sigma_j)_{j \neq i}$ in (2) above.

We want to show that for ICR it is only the players' beliefs and higher order beliefs about states of nature—their "Mertens-Zamir types"—that matter. For this we need to define, for each type t_i in a finite type space $\mathcal{T} = (T_i, \pi_i)_{i=1}^{I}$, that type's beliefs and higher order beliefs about Θ. Let

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

For each $k = 2, 3, ...$, let

$$\widehat{\pi}_i^k(t_i)\left[\left(\left(\widetilde{\pi}_j^1,...,\widetilde{\pi}_j^{k-1}\right)_{j\neq i},\theta\right)\right] = \pi_i(t_i)\left[\left\{(t_{-i},\theta):\left(\left(\widehat{\pi}_j^\kappa[t_j]\right)_{j\neq i}\right)_{\kappa=1}^{k-1} = \left(\left(\widetilde{\pi}_j^\kappa\right)_{j\neq i}\right)_{\kappa=1}^{k-1}\right\}\right].$$

Finally, let

$$\widehat{\pi}_i^*(t_i) = \left(\widehat{\pi}_i^k(t_i)\right)_{k=1}^{\infty}.$$

3 Properties of the solution concept

3.1 Dependence on types but not on type spaces

Proposition 1 *If t_i is a type in a finite type space $\mathcal{T}$, t'_i is a type in finite type space $\mathcal{T}'$ and $\widehat{\pi}_i^*(t_i) = \widehat{\pi}_i^*(t'_i)$, then $R_i^{\mathcal{T}}(t_i) = R_i^{\mathcal{T}}(t'_i)$.*

The proof is by induction. Intuitively, the first-level rationalizable actions R_1 are those that are best responses to arbitrary conjectures about the opponents; because conjectures allow correlation with θ, these depend only on beliefs about Θ. Second-level rationalizability depends on beliefs about Θ and the opponents' first-level rationalizable sets; these in turn depend only on the opponent's first-level beliefs, so second-level rationalizability is determined by second-level beliefs, and so on.

Proof. We establish by induction for each k that if $\widehat{\pi}_i^*(t_i) = \widehat{\pi}_i^*(t'_i)$ then $R_{i,k}^{\mathcal{T}}(t_i) = R_{i,k}^{\mathcal{T}'}(t'_i)$. Suppose that this holds for $k-1$, that $\widehat{\pi}_i^*(t_i) = \widehat{\pi}_i^*(t'_i)$ and that $a_i \in R_{i,k}^{\mathcal{T}}(t_i)$. Thus there exists $\nu \in \Delta(T_{-i}\times\Theta\times A_{-i})$ such that

$$\begin{aligned}
&(1)\ \nu[(t_{-i},\theta,a_{-i})] > 0 \Rightarrow a_{-i} \in \left(R_{j,k-1}^{\mathcal{T}}(t_j)\right)_{j\neq i}\\
&(2)\ \sum_{a_{-i}} \nu[(t_{-i},\theta,a_{-i})] = \pi_i(t_i)[(t_{-i},\theta)]\\
&(3)\ a_i \in \arg\max_{a'_i} \sum_{T_{-i}\times\Theta\times A_{-i}} g_i((a'_i,a_{-i}),\theta)\,\nu[(t_{-i},\theta,a_{-i})]
\end{aligned}$$

We now construct a $\nu' \in \Delta\left(T'_{-i}\times\Theta\times A_{-i}\right)$ such that the above three conditions hold when ν' and $\pi_i(t'_i)$ replace ν and $\pi_i(t_i)$, respectively. Let $D_{-i}^{k-1} = \left\{\left(\widehat{\pi}_j^{k-1}(t_{-i})\right)_{j\neq i} : t_{-i}\in T_{-i}\right\}$. For $\widetilde{\pi}_{-i}^k \in D_{-i}^{k-1}$, let $\gamma\left(\widetilde{\pi}_{-i}^k,\theta\right) = \sum_{\left\{(t_{-i},a_{-i}):\left(\widehat{\pi}_j^{k-1}(t_j)\right)_{j\neq i}=\widetilde{\pi}_{-i}^{k-1}\right\}} \nu[(t_{-i},\theta,a_{-i})]$. Then for $\left(\widetilde{\pi}_{-i}^k,\theta\right)$

such that $\gamma\left(\widetilde{\pi}_{-i}^{k},\theta\right)>0$, set

$$\sigma_i\left(\widetilde{\pi}_{-i}^{k-1},\theta\right)[a_{-i}]=\frac{\sum_{\left\{t_{-i}:\left(\widehat{\pi}_j^{k-1}(t_j)\right)_{j\neq i}=\widetilde{\pi}_{-i}^{k-1}\right\}}\nu\left[(t_{-i},\theta,a_{-i})\right]}{\gamma\left(\widetilde{\pi}_{-i}^{k},\theta\right)}$$

and for all other $\left(\widetilde{\pi}_{-i}^{k},\theta\right)$ set

$$\sigma_i\left(\widetilde{\pi}_{-i}^{k-1},\theta\right)[a_{-i}]=\left\{\begin{array}{c}1/\#\left(R_{j,k-1}^{\mathcal{T}}(t_j)\right)_{j\neq i}\text{ if }a_{-i}\in\left(R_{j,k-1}^{\mathcal{T}}(t_j)\right)_{j\neq i}\\ 0\text{ otherwise}\end{array}\right.$$

Next, let

$$\nu'\left[\left(t'_{-i},\theta,a_{-i}\right)\right]=\pi_i\left(t'_i\right)\left[\left(t'_{-i},\theta\right)\right]\sigma_i\left(\widehat{\pi}_{-i}^{k-1}\left(t'_{-i}\right),\theta\right)[a_{-i}],$$

where $\widehat{\pi}_{-i}^{k-1}\left(t'_{-i}\right)\in D_{-i}^{k-1}$ since $\widehat{\pi}_i^{*}(t_i)=\widehat{\pi}_i^{*}(t'_i)$.

Thus

$$\begin{aligned}\gamma\left(\widetilde{\pi}_{-i}^{k},\theta\right) &= \sum_{\left\{(t_{-i},a_{-i}):\left(\widehat{\pi}_j^{k-1}(t_j)\right)_{j\neq i}=\widetilde{\pi}_{-i}^{k-1}\right\}}\nu\left[(t_{-i},\theta,a_{-i})\right]\\ &= \pi_i(t_i)\left[\left\{t_{-i}:\left(\widehat{\pi}_j^{k-1}(t_j)\right)_{j\neq i}=\widetilde{\pi}_{-i}^{k-1}\right\}\times\{\theta\}\right]\\ &= \pi_i(t'_i)\left[\left\{t_{-i}:\left(\widehat{\pi}_j^{k-1}(t_j)\right)_{j\neq i}=\widetilde{\pi}_{-i}^{k-1}\right\}\times\{\theta\}\right]\end{aligned}$$

Hence we obtain the following.

$$\begin{aligned}\sum_{t_{-i}}\nu'\left[(t_{-i},\theta,a_{-i})\right] &= \sum_{t_{-i}}\pi_i(t'_i)\left[(t_{-i},\theta)\right]\sigma_i\left(\widehat{\pi}_{-i}^{k-1}(t_{-i}),\theta\right)[a_{-i}]\\ &= \sum_{\widetilde{\pi}_{-i}^{k-1}\in D_{-i}^{k-1}}\pi_i(t'_i)\left[\left\{t_{-i}:\left(\widehat{\pi}_j^{k-1}(t_j)\right)_{j\neq i}=\widetilde{\pi}_{-i}^{k-1}\right\}\right]\sigma_i\left(\widehat{\pi}_{-i}^{k-1}(t_{-i}),\theta\right)[a_{-i}]\\ &= \sum_{\widetilde{\pi}_{-i}^{k-1}\in D_{-i}^{k-1}}\gamma\left(\widetilde{\pi}_{-i}^{k},\theta\right)\frac{\sum_{\left\{t_{-i}:\left(\widehat{\pi}_j^{k-1}(t_j)\right)_{j\neq i}=\widetilde{\pi}_{-i}^{k-1}\right\}}\nu\left[(t_{-i},\theta,a_{-i})\right]}{\gamma\left(\widetilde{\pi}_{-i}^{k},\theta\right)}\\ &= \sum_{t_{-i}}\nu\left[(t_{-i},\theta,a_{-i})\right] \qquad (1)\end{aligned}$$

So ν and ν' have the same marginal distributions on $A_{-i}\times\Theta$.

Now we claim

$$\begin{aligned}
&(1)\ \nu'\left[(t_{-i},\theta,a_{-i})\right]>0\Rightarrow a_{-i}\in\left(R^{\mathcal{T}}_{j,k-1}\left(t_j\right)\right)_{j\neq i}\\
&(2)\ \sum_{a_{-i}}\nu'\left[(t_{-i},\theta,a_{-i})\right]=\pi_i\left(t'\right)\left[(t_{-i},\theta)\right]\\
&(3)\ a_i\in\arg\max_{a_i'}\sum_{T_{-i}\times\Theta\times A_{-i}}g_i\left(\left(a_i',a_{-i}\right),\theta\right)\nu'\left[(t_{-i},a_{-i},\theta)\right]
\end{aligned}$$

(1) is true by the inductive hypothesis and the construction, (2) by construction, and (3) because ν and ν' have the same marginal distributions on $A_{-i}\times\Theta$. So $a_i'\in R^{\mathcal{T}}_{i,k}\left(t_i'\right)$. ■

Note that our construction of σ_i in the proof required that conjectures about actions be correlated with θ.

3.2 Discussion

As we have pointed out, interim correlated rationalizability allows for conjectures about the state and a type's actions to be correlated; this correlation is what generates the result that the solution concept depends only on a type, and not details of the type space. Such correlation can have surprising consequences, as in the following example.

Example 2 *There are two states of Nature and it is common knowledge that each player assigns probability* $1/2$ *to each state. Each player decides whether to bet (action* B*) or not (action* N*). If both players chose* B*, they transfer* 3 *or* -3 *from one to the other depending on the state, and choosing* B *incurs a cost of 1 regardless of the opponent's action. This generates the following payoff functions:*

θ	B	N
B	$2,-4$	$-1,0$
N	$0,-1$	$0,0$

θ'	B	N
B	$-4,2$	$-1,0$
N	$0,-1$	$0,0$

In this game, it is ICR for each player to choose B *(where* i *believes* j *chooses* B *only when the state is favorable to* i*), as* B *is a best reply against the belief that assigns probability* $1/2$ *to* (θ,B) *and* $1/2$ *to* (θ',N) *for the row player (and with the opposite beliefs for the column player); this belief has the correct marginal on* Θ *as required.*

Thus each player expects there to be costly speculative trade (and indeed using the epistemic set-up of the next section, trade can be seen to be common knowledge) even though there *is a common prior. This possibility relies on each player believing in correlation between the other player's actions and the state.*[9]

To justify the original definition of independent rationalizability in Bernheim (1984) and Pearce (1984), it is necessary to add additional conditional independence assumptions. The question of whether or not to impose the assumptions parallels an older debate in the complete-information environment. Brandenburger and Dekel (1987) showed that correlated rationalizability (allowing players to have correlated conjectures over others' actions) corresponds to common knowledge of rationality. To interpret this correlation, it is important to remember that a player's conjectures represent his subjective beliefs about the distribution of play; any correlations in these beliefs need not correspond to "objective correlation" that would be seen by an outside observer. The correlations we consider in this paper should be interpreted in the same way.

There has been increasing acceptance of using the correlated version of rationalizability, in part based on the influential argument of Aumann (1987):

> ...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 has no connection with any overt or even covert collusion between 1 and 2; they may be acting entirely independently....

Interim correlated rationalizability extends this view, by 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?

---

[9] Although there is a common prior over Θ, this observation is not inconsistent with no-trade theorems because there is not common knowledge of the conditional probability of trade in each state θ. This lack of common knowledge is possible because ICR allows beliefs about strategic behavior that are not consistent with a common prior, and this, as in complete-information games, allows each player to think that he is "outguessing" the other. Note that if we set the payoff to choosing action B when the opponent chooses N to be -4 instead of -1, then action B is no longer rationalizable for any common-prior type, although it remains rationalizable for some non-common-prior types. Thus common-prior and non-common-prior types can be distinguished in some no-trade games. Dekel, Fudenberg and Morris (2006) use this observation to show that finite common-prior types are not dense in the universal type space in the strategic topology that they define.

One might argue that any correlation—about players or about Nature—should be made explicit. We take the opposing, "small-worlds," view that such correlation may not be an inherent part of the interaction being studied, and hence is best incorporated into the solution concept and not the model.[10]

Moreover, we view incomplete-information games as a tool to study an interim environment. With this viewpoint; the ex-ante type space is an artificial construct, and it seems difficult to know which is the right ex-ante type space to impose. Therefore, it seems best to adopt a solution concept that does not depend on which variant is adopted. Finally, as we explain in more detail in subsection 3.4 below, if one embeds a given type space into an epistemic model, common knowledge of rationality (without further assumptions on independence of beliefs or common priors) corresponds to ICR.

One might also argue in favor of a hybrid solution concept—in between ICR and IIR—that allows arbitrary correlation in conjectures about other players but insists that the correlation with Nature is explicitly captured in the type space.[11] A difficulty with such hybrid notions is that the resulting solution concept will be sensitive to the addition of a dummy player who any single other player believes is omniscient. That is, the existence of a player k whom i thinks knows more about the state of Nature than does j, enables i to believe j's actions are correlated with Nature via k. Hence, if one allows for correlation with opponents but not with Nature then games must completely specify all agents, even if their actions are not payoff relevant.

Example 3 (Example 2 continued) *In the preceding betting game the only IIR action is* N*. Now add a third player to the game who chooses an action* $a_3 \in A_3 = \{B, N\}$*. The payoffs to players 1 and 2 are unchanged, and unaffected by player 3's choice, while player 3 payoff is constant. Player* 3 *has two possible types* t_3 *and* t_3' *who know whether the state is* θ *or* θ'*. IIR requires independence across opponents and nature, and hence this has no effect. However,if one allows arbitrary correlations in beliefs about players actions but not*

---

[10] Of course, if one wants to explicitly model and study the effect of different forms of correlation one would need to use a different solution concept (such as IIR) that does not implicitly allow all such correlation.

[11] Ely and Peski (2005) study a definition of interim rationalizability in two-player incomplete-information games that is equivalent to our definition of interim independent rationalizability (in two-player games). They suggest that, in many-player games, one might want to examine hybrid notions of interim rationalizability such as the one we criticize here.

with Nature, then the resulting "interim hybrid rationalizability" solution concept (which we have not formally defined) would allow for B (as well as N), as player 1 could believe that player 2's play is correlated with player 3's, and that player 3's play is correlated with θ.[12]

3.3 Equivalent formulations

We provide some obvious equivalent definitions that further illustrate the analogies to the complete-information environment.

3.3.1 Iterated Undominance

Fudenberg and Tirole (1991) demonstrate the important distinction between interim and ex ante (strictly) dominated strategies. As one would expect, it is the interim version that is related to interim rationalizability. Specifically, iteratively deleting strategies that are not interim best replies is equivalent to iterated deletion of strictly interim dominated strategies (where beliefs in both are allowed to be correlated). Let $U^{\mathcal{T}}_{i,0}(t_i) = A_i$;

$$U^{\mathcal{T}}_{i,k+1}(t_i) = \left\{ a_i \in A_i \;\middle|\; \begin{array}{l} \text{there does not exist } \alpha_i \in \Delta(A_i) \text{ such that} \\ \displaystyle\sum_{T_{-i}\times\Theta\times A_{-i}} g_i\left((a_i, a_{-i}), \theta\right) \nu\left[(t_{-i}, \theta, a_{-i})\right] < \\ \displaystyle\sum_{A_i}\sum_{T_{-i}\times\Theta\times A_{-i}} \alpha_i(a_i') g_i\left((a_i', a_{-i}), \theta\right) \nu\left[(t_{-i}, \theta, a_{-i})\right] \\ \text{ for all } \nu \in \Delta(T_{-i}\times\Theta\times A_{-i}) \text{ such that} \\ (1)\ \nu\left[(t_{-i}, \theta, a_{-i})\right] > 0 \Rightarrow a_{-i} \in \left(U^{\mathcal{T}}_{j,k}(t_j)\right)_{j\neq i} \\ (2)\ \displaystyle\sum_{A_{-i}} \nu\left[(t_{-i}, \theta, a_{-i})\right] = \pi_i(t_i)\left[(t_{-i}, \theta)\right] \end{array} \right\};$$

and

$$U^{\mathcal{T}}_i(t_i) = \lim_{k\to\infty} U^{\mathcal{T}}_{i,k}(t_i).$$

Claim 1 *$R^{\mathcal{T}}_i(t_i) = U^{\mathcal{T}}_i(t_i)$.*

---

[12] The conclusion about the effect of dummy players also holds in a model where player 3 has a third possible type t_3'' and player 2 is certain that player 3 is t_3'': what is important is only that player 1 believes that player 3 knows θ. Unlike the example in the text, this version does not reproduce the entire set of ICR actions.

3.3.2 Best-reply sets

Similarly, there is an obviously equivalent best-reply set (Pearce (1984)) definition of ICR. Let $S_i^{\mathcal{T}} : T_i \to 2^{A_i} \big/ \varnothing$ be a specification of possible actions for each type, and $S^{\mathcal{T}} = \left(S_1^{\mathcal{T}}, ..., S_I^{\mathcal{T}}\right)$.[13]

Definition 1 *$S^{\mathcal{T}}$ is a best-reply set if for each t_i and $a_i \in S_i^{\mathcal{T}}(t_i)$, there exists $\sigma_{-i} : T_{-i} \times \Theta \to \Delta(A_{-i})$ such that*

$$\begin{array}{ll} (1) & \sigma_{-i}(t_{-i}, \theta)[a_{-i}] > 0 \Rightarrow a_{-i} \in S_{-i}^{\mathcal{T}}(t_{-i}) \\ (2) & a_i \in \arg\max_{a_i'} \sum_{t_{-i}, a_{-i}, \theta} \pi(t_i)[(t_{-i}, \theta)]\, \sigma_{-i}(t_{-i}, \theta)[a_{-i}]\, g_i((a_i', a_{-i}), \theta) \end{array}$$

Claim 2 *1. If $S_c^{\mathcal{T}}$ for all c in some index set C are best-reply sets then $\cup_c S_c^{\mathcal{T}}$ is a best-reply set.*

2. The union of all best-reply sets is equal to $\left(\left(R_i^{\mathcal{T}}(t_i)\right)_{t_i \in T_i}\right)_{i \in I}$.

Property 1 is immediate. That the union includes $R_i^{\mathcal{T}}$ follows from that observation that $\left(\left(R_i^{\mathcal{T}}(t_i)\right)_{t_i \in T_i}\right)_{i \in I}$ is a best-reply set. To see the converse, note that no action in a best-reply set can be deleted at any stage of the iteration, since at each point in the iteration each such action is a best reply to actions in the best-reply set, and hence remains.

3.3.3 Fixed points of a best-reply correspondence

Lastly we provide a fixed-point definition of $R_i^{\mathcal{T}}$. The best-reply correspondence takes as given a feasible subset of actions for each type of each opponent of i, and for each type t_i of i, determines the set of best replies.

Definition 2 *The correspondence of best replies for all types given a subset of actions for all types is denoted $BR^{\mathcal{T}} : \left(\left(2^{A_i}\right)_{t_i \in T_i}\right)_{i \in I} \to \left(\left(2^{A_i}\right)_{t_i \in T_i}\right)_{i \in I}$ and is defined as follows. First, given given $F = \left((F_{t_i})_{t_i \in T_i}\right)_{i \in I} \in \left(\left(2^{A_i}\right)_{t_i \in T_i}\right)_{i \in I}$ the best replies for t_i are*

$$BR_i^{\mathcal{T}}(t_i, F) = \left\{ a_i \in A_i \;\middle|\; \begin{array}{l} \text{there exists } \sigma_{-i} : T_{-i} \times \Theta \to \Delta(A_{-i}) \text{ such that} \\ (1)\ \sigma_{-i}(t_{-i}, \theta)[a_{-i}] > 0 \Rightarrow a_{-i} \in F_{t_{-i}} \\ (2)\ a_i \in \arg\max_{a_i'} \sum_{T_{-i} \times \Theta \times A_{-i}} g_i((a_i', a_{-i}), \theta)\, \sigma_{-i}(t_{-i}, \theta)[a_{-i}]\, \pi(t_i)[(t_{-i}, \theta)] \end{array} \right\}$$

---

[13] We abuse notation by calling $S^{\mathcal{T}}$ a "set" to emphasize the link to the complete information case; it is a correspondence.

Next we define [14]

$$BR^{\mathcal{T}}(F) = \left(\left(BR_i^{\mathcal{T}}(t_i, F) \right)_{t_i \in T_i} \right)_{i \in I}.$$

Claim 3 *The largest fixed point of* $BR^{\mathcal{T}}$ *is* $\left(\left(R_i^{\mathcal{T}}(t_i) \right)_{t_i \in T_i} \right)_{i \in I}$.

This follows from any fixed point corresponding to a best-reply set and the previous claim regarding best-reply sets.

Remark 1 *The definition of the best-reply correspondence is obviously equivalent to the following.*

$$BR_i^{\mathcal{T}}(t_i, F) = \left\{ a_i \in A_i \;\middle|\; \begin{array}{l} \text{there exists } \nu \in \Delta(T_{-i} \times \Theta \times A_{-i}) \text{ such that} \\ \text{(1) } \nu\left[\left\{ (t_{-i}, \theta, a_{-i}) : a_{-i} \in F_{t_{-i}} \right\} \right] = 1 \\ \text{(2) } a_i \in \arg\max\limits_{a_i'} \sum\limits_{T_{-i} \times \Theta \times A_{-i}} g_i\left((a_i', a_{-i}), \theta \right) \nu[(t_{-i}, \theta, a_{-i})] \\ \text{(3) } \left(marg_{T^*_{-i} \times \Theta} \nu \right) [(t_{-i}, \theta)] = \pi(t_i)[(t_{-i}, \theta)] \sum\limits_{A_{-i}} \nu[(t_{-i}, \theta, a_{-i})] \end{array} \right\}$$

A rewriting similar to this is used in proofs in the case of infinite types but as it is easier to understand in the finite types case, we present the finite form here.

3.4 Epistemic Foundations

To better understand the two solution concepts, ICR and IIR, we relate them to common knowledge of rationality. In order to do this, we introduce a richer language—an "epistemic model"—to model the knowledge of the players. We are able to provide an epistemic foundation for the solution concepts in the spirit of the existing epistemic-foundations literature.[15] We note that—at least in our epistemic formulation—additional common-knowledge assumptions are necessary to provide an epistemic foundation for interim independent rationalizability. We discuss this further at the end of this subsection.

Throughout this section, we fix a type space $\mathcal{T} = (T_i, \pi_i)_{i=1}^{I}$. We will sometimes refer to this object as a "standard" type space and to elements of T_i as standard types. As

---

[14] We abuse notation and write BR both for the correspondence specifying best replies for a type and for the correspondence specifying these actions for all types.

[15] Aumann (1987), Brandenburger and Dekel (1987), Tan and Werlang (1988), Aumann and Brandenburger (1995).

discussed, we view this type space as exogenously given, and not necessarily a complete description (in particular not necessarily including all possible correlation). We then assume only that this type space (and the game and rationality) is common knowledge. That is, we embed the basic type space in an arbitrary larger space, the epistemic space—which can be any extension to a more description of the players' perception of the world, specifying at least their actions at any state—and we assume that the (original) type space is common knowledge in this epistemic space. Then we ask what can we say about play in game defined by the original type space; i.e., what solution concept—defined on games with the original type space—is characterized.

Let E_i be a finite set of epistemic types for player i, and let $E = (E_i)_{i \in I}$. An epistemic model specifies for each i how e_i determines

1. beliefs over the types of others and the payoff states, $\phi_i : E_i \to \Delta(E_{-i} \times \Theta)$;
2. i's action,
3. **i's** "standard type," $\tau_i : E_i \to T_i$**.**

Thus an epistemic model consists of $(E_i, \phi_i, a_i^*, \tau_i)_{i=1}^{I}$; its state space is $\Omega = E \times \Theta$.

There are some events in which we are particularly interested. For a given epistemic model, we write Rat_i for the set of states where player i is "rational",

$$Rat_i = \left\{ \left(\left(e_i, e'_{-i}\right), \theta'\right) \middle| a_i^*\left(e_i\right) \in \arg\max_{a_i} \sum_{E_{-i} \times \Theta} g_i\left(\left(a_i, a_{-i}^*\left(e_{-i}\right)\right), \theta\right) \phi_i\left(e_i\right)\left[\left(e_{-i}, \theta\right)\right] \right\},$$

and Rat for the set of states where all players are rational,

$$Rat = \bigcap_i Rat_i.$$

We write W_i for the set of states where player i has the correct beliefs about $T_{-i} \times \Theta$ given his type

$$W_i = \left\{ \left(\left(e_i, e'_{-i}\right), \theta'\right) \middle| \sum_{\left\{e_{-i} : \left(\tau_j\left(e_j\right)\right)_{j \neq i} = t_{-i}\right\}} \phi_i\left(e_i\right)\left[\left(e_{-i}, \theta\right)\right] = \pi_i\left(\tau_i\left(e_i\right)\right)\left[\left(t_{-i}, \theta\right)\right] \right\},$$

$$W = \bigcap_i W_i.$$

The idea here is that there is some type space (it would be part of what is in the minds of the players, or it could model an objective ex ante information structure), and we will want to assume that the beliefs in the type space are common knowledge in the epistemic space.

The set of states where individual i knows the event $H \subseteq \Omega$ is

$$K_i(H) = \left\{ \left(\left(e_i, e'_{-i}\right), \theta'\right) \middle| \sum_{\{(e_{-i},\theta):((e_i,e_{-i}),\theta)\in H\}} \phi_i(e_i)\left[(e_{-i},\theta)\right] = 1 \right\},$$

the set of states where everyone knows the event H is

$$K_*(H) = \underset{i}{\cap} K_i(H),$$

and the set of states where there is common knowledge of H is

$$CK(H) = \underset{n}{\cap} (K_*)^n (H).$$

Proposition 2 *Interim Correlated Rationalizability characterizes common knowledge of rationality and of the standard type space, i.e.,*

1. *in any epistemic model, if* $((e_i, e_{-i}), \theta) \in CK(Rat \cap W)$*, then* $a_i^*(e_i) \in R_i^{\mathcal{T}}(\tau_i(e_i))$*;*
2. *There is an epistemic model such that if if* $a_i \in R_i^{\mathcal{T}}(t_i)$*, then there is a state* $((e_i, e_{-i}), \theta)$ *such that* $((e_i, e_{-i}), \theta) \in CK(Rat \cap W)$*,* $\tau_i(e_i) = t_i$ *and* $a_i^*(e_i) = a_i$

Proof. (1) Suppose $\left(\left(e_i^*, e_{-i}^*\right), \theta^*\right) \in CK(Rat)$. Let E_j^* be the set of epistemic types of player j where j knows $CK(Rat)$. Let

$$S_i(t_i) = \left\{a_i \mid \text{for some } e_i \in E_i^*,\ a_i^*(e_i) = a_i \text{ and } \tau_i(e_i) = t_i\right\}.$$

Observe that, by construction,

$$a_i^*(e_i^*) \in S_i(\tau_i(e_i^*)).$$

Now for any $a_i \in S_i(t_i)$, pick any $e_i \in E_i^*$ such that $a_i^*(e_i) = a_i$ and $\tau_i(e_i) = t_i$. Let

$$\lambda_i^{a_i,t_i}\left[(t_{-i}, \theta, a_{-i})\right] = \sum_{\left\{(e_{-i},\theta):\tau_{-i}(e_{-i})=t_{-i} \text{ and } a_{-i}^*(e_{-i})=a_{-i}\right\}} \phi_i(e_i)\left[(e_{-i}, \theta)\right].$$

Again by construction,

$$a_i \in \arg\max_{a_i'} \sum_{T_{-i}\times\Theta\times A_{-i}} \lambda_i^{a_i,t_i}\left[(t_{-i},\theta,a_{-i})\right] g_i\left(\left(a_i',a_{-i}\right),\theta\right).$$

Common knowledge of W ensures that $\sum\limits_{a_{-i}} \lambda_i^{a_i,t_i}\left[(t_{-i},\theta,a_{-i})\right] = \pi_i\left(t_i\right)\left[(t_{-i},\theta)\right]$ for all t_{-i},θ. Thus an inductive argument ensures that $S_i\left(t_i\right) \subseteq R_{i,k}^{\mathcal{T}}\left(t_i\right)$ for all k and thus $S_i\left(t_i\right) \subseteq R_i^{\mathcal{T}}\left(t_i\right)$. So

$$a_i^*\left(e_i^*\right) \in S_i\left(\tau_i\left(e_i^*\right)\right) \subseteq R_i^{\mathcal{T}}\left(\tau_i\left(e_i^*\right)\right).$$

(2) To begin we construct the epistemic spaces "pointiwise," i.e. we construct a seperate epsitemic spaces for every action a_i that is an element of $R_i^{\mathcal{T}}\left(t_i^*\right)$ for some t_i^*. Then we explain how to combine these into one larger epsitemic space that has the stated properties for all such actions. So fix a player i, a type t_i^*, and suppose that $a_i^* \in R_i^{\mathcal{T}}\left(t_i^*\right)$. We will construct an epistemic type space. Let $E_i = \left\{(t_i,a_i) : a_i \in R_i^{\mathcal{T}}\left(t_i\right)\right\}$. Let

$$\begin{aligned} a_i^*\left(e_i\right) &= a_i^*\left(\left(t_i,a_i\right)\right) = a_i \\ \tau_i\left(e_i\right) &= \tau_i\left(\left(t_i,a_i\right)\right) = t_i. \end{aligned}$$

Observe that for each $a_i \in R_i^{\mathcal{T}}\left(t_i\right)$, there exists $\lambda_i^{a_i,t_i} \in \Delta\left(T_{-i}\times\Theta\times A_{-i}\right)$ such that

$$\left.\begin{array}{l} (1)\ \lambda_i^{a_i,t_i}\left[(t_{-i},\theta,a_{-i})\right] > 0 \Rightarrow a_{-i} \in \left(R_j^{\mathcal{T}}\left(t_j\right)\right)_{j\neq i} \\ (2)\ \sum\limits_{A_{-i}} \lambda_i^{a_i,t_i}\left[(t_{-i},\theta,a_{-i})\right] = \pi_i\left(t_i\right)\left[(t_{-i},\theta)\right] \text{ for all } t_{-i},\theta \\ (3)\ a_i \in \arg\max\limits_{a_i'} \sum\limits_{T_{-i}\times\Theta\times A_{-i}} \lambda_i^{a_i,t_i}\left[(t_{-i},\theta,a_{-i})\right] g_i\left(\left(a_i',a_{-i}\right),\theta\right) \end{array}\right. ;$$

Let

$$\phi_i\left(e_i\right)\left[(e_{-i},\theta)\right] = \phi_i\left(t_i,a_i\right)\left[\left(\left(t_j,a_j\right)_{j\neq i},\theta\right)\right] = \lambda_i^{a_i,t_i}\left[(t_{-i},\theta,a_{-i})\right].$$

We have now constructed an epistemic type space that embeds the standard type space, with $Rat_i = \Omega$ for all i and thus $\left(\left(\left(t_i^*,a_i^*\right),e_{-i}\right),\theta\right) \in CK\left(Rat\right)$ for all e_{-i},θ. This shows that for each action there is an epsitemic type space as desired.

To conclude it is trivial to combine these epistemic type spaces together into one large epistemic type space. Note that given two epistemc models, $\left(E_i^1,\phi_i^1,a_i^{*1},\tau_i^1\right)_{i=1}^I$ and $\left(E_i^2,\phi_i^2,a_i^{*2},\tau_i^2\right)_{i=1}^I$ they can be combined in the obvious way to obtain a "joint" epistemic model $\left(\tilde{E}_i,\tilde{\phi}_i,\tilde{a}_i^*,\tilde{\tau}_i\right)_{i=1}^I$, with state space $\tilde{\Omega} = \tilde{E}\times\Theta$, where $\tilde{E}_i = E_i^1 \cup E_i^2$, $\tilde{\phi}_i\left(\tilde{e}_i\right)\left[\tilde{e}_{-i},\theta\right] = \phi_i^m\left(e_i^m\right)\left[e_{-i}^m,\theta\right]$ if

$\tilde{e} = e^m \in E^m$, and zero otherwise, $\tilde{a}_i^* (\tilde{e}_i) = \tilde{a}_i^{*m} (e_i^m)$ and $\tilde{\tau}_i (\tilde{e}_i) = \tau_i^m (e_i^m)$ if $\tilde{e}_i = e_i^m$. Moreover, if an event in $E^m \times \Theta$ is common knowledge at state (e^m, θ) then the same event is an event in $\tilde{\Omega}$ and it is common knowledge at $(\tilde{e}, \theta) = (e^m, \theta)$. ■

Remark 2 *A standard reinterpretation of the result is that if we start with the standard type space* $\mathcal{T} = (T_i, \pi_i)_{i=1}^I$, *we can construct a larger type space and a morphism from the original type space to the larger type space which preserves all properties of the standard type space, and a Nash equilibrium on that larger type space, such that for each type in the original type space and each interim correlated rationalizable action for that type, there is a corresponding type in the larger space who plays that action in equilibrium.*

Remark 3 *A referee noted that one could give a different interpretation of our epistemic analysis: At states in the epistemic type space where there is common knowledge of rationality, every player will be choosing an interim independent rationalizable action for his type in that epistemic type space, so there is a sense in which one can interpret the result as yielding IIR and not ICR. To understand the difference between these two interpretations, consider Aumann's (1987) characterization of correlated equilibrium. Aumann essentially assumed a singleton type space (i.e. a complete information game) that was embedded in an epistemic space, and then showed that if there is a common prior on the epistemic space then the distribution of actions corresponds to a correlated equilibrium distribution on the original game. Adopting the alternative approach of the referee would imply asking what is played on the enlarged game with the epistemic type space. This does not yield all the correlated equilibria; in two player games, it yields exactly the Nash equilibria. We follow Aumann in studying the implications of common-knowledge assumptions on any epistemic space in which a certain game (with a degenerate type space in Aumann's case, or a general type space in ours) and rationality of the players is common knowledge.*

Finally, we briefly note for comparison an epistemic characterization of interim independent rationalizability in our language. The set of states where player i has independent beliefs, i.e., believes that each other player's type is a sufficient statistic for his behavior, is

$$Y_i = \left\{ \left(\left(e_i, e'_{-i}\right), \theta'\right) \left| \begin{array}{c} \text{for each } j \neq i, \text{ there exists } \sigma_j : T_j \longrightarrow A_j \text{ such that} \\ \displaystyle\sum_{\left\{(e_{-i},\theta):\tau_{-i}(e_{-i})=t_{-i} \text{ and } a^*_{-i}(e_{-i})=a_{-i}\right\}} \phi_i\left(e_i\right)\left[\left(e_{-i},\theta\right)\right] \\ = \left(\displaystyle\sum_{\left\{(e_{-i},\theta):\tau_{-i}(e_{-i})=t_{-i}\right\}} \phi_i\left(e_i\right)\left[\left(e_{-i},\theta\right)\right] \right) \displaystyle\prod_{j\neq i} \sigma_j\left(t_j\right)\left[a_j\right] \end{array} \right. \right\}$$

$$Y = \bigcap_i Y_i$$

Proposition 3 *Independent Interim Rationalizability characterizes common knowledge of rationality, the standard type space and independent beliefs, i.e.,*

1. *in any epistemic model, if* $\left(\left(e_i, e_{-i}\right), \theta\right) \in CK\left(Rat \cap W \cap Y\right)$*, then* $a^*_i\left(e_i\right) \in IIR^*_i\left(\tau_i\left(e_i\right)\right)$*;*
2. *if* $a_i \in IIR^*_i\left(t_i\right)$*, then there exists an epistemic model and a state* $\left(\left(e_i, e_{-i}\right), \theta\right)$ *such that* $\left(\left(e_i, e_{-i}\right), \theta\right) \in CK\left(Rat \cap W \cap Y\right)$*,* $\tau_i\left(e_i\right) = t_i$ *and* $a^*_i\left(e_i\right) = a_i$*.*

The proof closely follows the proof of proposition 2 and hence is not provided. This result is the incomplete-information analog of Proposition 3.1 in Brandenburger and Dekel (1987).

This proposition shows that additional assumptions—beyond common knowledge of rationality and the type space—are need to justify restricting attention to actions that are interim independent rationalizable on the type space. The additional assumption of common knowledge of independent beliefs makes explicit the key idea underlying the solution concept: no unexplained correlation in beliefs is allowed.

4 Infinite Type Spaces

4.1 The type spaces

We now extend our analysis to type spaces that are not necessarily finite. To do so, we base our development on Heifetz and Samet's (1998) topology-free construction.

For measurable X we denote by $\Delta\left(X\right)$ the set of (probability) measures on X; all product spaces are endowed with the product sigma-algebra. The primitives of our model are a finite set Θ of states of Nature, a finite set $I = \{1, ..., I\}$ of players, and a type space $\mathcal{T} = \left(T_i, \pi_i\right)_{i=1}^{I}$.

We assume that each T_i is a measurable space, set $T_{-i} = \times_{j \neq i} T_j$, and give $T_{-i} \times \Theta$ the product sigma-algebra.[16]

Following Heifetz and Samet, we assume that for every measurable space X, the set $\Delta(X)$ of measures on X is endowed with the sigma-algebra generated by

$$\{\{\mu : \mu(Z) \geq p\} : p \in [0,1] \text{ and } Z \text{ a measurable subset of } X\}. \tag{2}$$

Each $\Delta(T_{-i} \times \Theta)$ gets the corresponding sigma algebra; we then assume that each $\pi_i : T_i \rightarrow \Delta(T_{-i} \times \Theta)$ is a measurable function. 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 $\pi_i(t_i)$ about the joint distribution of the opponent's type and the state of Nature. The above setup defines what Heifetz and Samet call a measurable type space.[17]

There is a belief-preserving morphism from one measurable type space into another measurable type space if it can be mapped into that space while preserving the belief structure. Formally, there is a belief-preserving morphism from (T_i, π_i) into $\left(\tilde{T}_i, \tilde{\pi}_i\right)$ if for each i there exists measurable $\varphi_i : T_i \rightarrow \tilde{T}_i$ with

$$\tilde{\pi}_i(\varphi_i(t_i))[Z] = \pi_i(t_i)\left[\left\{(t_{-i}, \theta) : \left(\theta, \varphi_{-i}(t_{-i})\right) \in Z\right\}\right]$$

for all measurable $Z \subset \tilde{T}_{-i} \times \Theta$, we call $\varphi = (\varphi_1, ..., \varphi_n)$ the morphism.

A particularly useful type space is a "universal type space" that we describe next. 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 algebra described above, and each X_k is given the product algebra over its two components. An element $(\delta_1, \delta_2, ...) \in (\Delta(X_k))_{k=0}^{\infty} \triangleq H$ 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, $T_i^* \subset H$, and a measurable belief function, $\pi_i^* : T_i^* \rightarrow \Delta\left(T_{-i}^* \times \Theta\right)$, for all i. Note that since there is a common uncertainty space Θ, the sets T_i^* are copies of the same set T^*. Therefore, where no confusion results, we drop the subscript of i for notational simplicity. The type space is

---

[16] The measurable structure on player i's beliefs is used to model the beliefs of other players about i's type. The set-up here, which is standard, implicitly assumes that any two players i and j have the same measurable structure on the types of a third player k

[17] Heifetz and Samet allowed Θ to be a general measurable space. We continue to endow Θ and all other finite sets with the obvious σ-algebra.

universal in that there is a unique belief-preserving morphism of any other measurable type space 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 (measurable) set of k^{th}-level beliefs for all types in T^*, $T_k^* \subseteq \Delta (X_{k-1})$. Given any measurable type space type t_i's marginal beliefs about the state of Nature are

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

For each $k = 2, 3..$ and measurable $Z \subseteq X_{k-1}$, let

$$\widehat{\pi}_i^k (t_i) [Z] = \pi_i (t_i) \left[\left\{ t_{-i} \in T_{-i} \,\middle|\, \left(\widehat{\pi}_{-i}^1 (t_{-i}) , ..., \widehat{\pi}_{-i}^{k-1} (t_{-i}) , \theta \right) \in Z \right\} \right] ;$$

the morphism guarantees that $\widehat{\pi}_i^k : T_i \rightarrow T_k^*$ is measurable for each k.

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

The connection between this and the topological construction of the universal type space (Mertens and Zamir (1985); see also Brandenburger and Dekel (1993), Heifetz (1993), Mertens, Sorin and Zamir (1994)) is clarified by the following claims, which are standard; see, e.g., footnote 5 in Battigalli and Siniscalchi (1999) and Ely and Peski (2005). The first part of the claim states that the sigma-algebra defined by Heifetz and Samet (1998) is the Borel algebra for the weak topology when the domain is Polish. Hence we also use $\Delta (X)$ to denote the Borel algebra when it is equivalent. This relationship implies the second part: that the hierarchies of beliefs constructed above coincide with the hierarchies that are constructed in Brandenburger and Dekel (1993) when Θ is Polish (a fortiori finite). Since the universal type space is unique (Heifetz and Samet (1998, Proposition 3.5)), the topological construction of the universal type space and the measure theoretic construction result in the same universal type space (up to belief-preserving morphisms).

Claim 4 1. *For a Polish space X, the sigma-algebra on $\Delta(X)$ generated by Heifetz and Samet (1998) when X is endowed with the Borel algebra, is the same as the Borel algebra of $\Delta (X)$ when $\Delta (X)$ is endowed with the weak topology.*

2. *For finite Θ there is a belief-preserving isomorphism between the measure-theoretic universal type space constructed in Heifetz and Samet (1998) and the earlier topological constructions referred to above.*

4.2 Interim Correlated Rationalizability

We now restate some of our earlier definitions and prove for this environment the key analogous results. In many cases, the only changes the definitions require are easy to identify: sums need to be replaced by integrals, measurability conditions must be imposed, and finite probabilities must be replaced by measures. We describe in detail those few cases where extra care is required in the notation; for brevity, we do not repeat the definitions whose extensions are obvious.

4.2.1 Best replies

For any subset of actions for all types, we first define the best replies when beliefs over opponents' strategies are restricted to those actions. We write

$$\int\limits_{T_{-i}\times\Theta\times A_{-i}} f(\cdot)\,\nu(\cdot)\left[\mathrm{d}\left(t_{-i},\theta,a_{-i}\right)\right]$$

for the integral of a measurable function f on $T_{-i}\times\Theta\times A_{-i}$ under measure ν, and

$$\int\limits_{T_{-i}} f(\cdot)\,\nu(\cdot)\left[\left(\mathrm{d}t_{-i},\theta,a_{-i}\right)\right]$$

when integrating with respect to t_{-i} only, while holding θ and a_{-i} fixed .

Definition 3 *The correspondence of best replies for all types given a subset of actions for all types is denoted* $BR^{\mathcal{T}}:\left(\left(2^{A_i}\right)^{T_i}\right)_{i\in I}\to\left(\left(2^{A_i}\right)^{T_i}\right)_{i\in I}$ *and is defined as follows. First, given a specification of a subset of actions for each possible type ,* $F=\left(\left(F_{t_j}\right)_{t_j\in T_j}\right)_{j\in I}$, *with* $F_{t_j}\subset A_j$ *for all* t_j *and* $j\in I$, *we define the best replies for* t_i *as*

$$BR_i^{\mathcal{T}}\left(t_i,F\right)=\left\{a_i\in A_i\;\middle|\;\begin{array}{l}\text{there exists a measurable }\sigma_{-i}:T_{-i}\times\Theta\to\Delta\left(A_{-i}\right)\text{ such that}\\ \text{(1) }\sigma_{-i}\left(t_{-i},\theta\right)\left[a_{-i}\right]>0\Rightarrow a_j\in F_{t_j}\\ \text{(2) }a_i\in\arg\max\limits_{a_i'}\sum\limits_{\Theta\times A_{-i}}\int\limits_{T_{-i}}g_i\left(a_i',a_{-i},\theta\right)\sigma_{-i}\left(t_{-i},\theta\right)\left[a_{-i}\right]\pi\left(t_i\right)\left[\left(dt_{-i},\theta\right)\right]\end{array}\right\}\tag{3}$$

Next we define[18]

$$BR^{\mathcal{T}}(F)=\left(\left(BR_i^{\mathcal{T}}\left(t_i,F\right)\right)_{t_i\in T_i}\right)_{i\in I}.$$

---

[18] We abuse notation and write BR both for the correspondence specifying best replies for a type and for the correspondence specifying these actions for all types.

Remark 4 *Because $A_{-i} \times \Theta$ is finite, and utility depends only on actions and beliefs, the set of best responses given some F, $BR_i^{\mathcal{T}}(t_i, F)$, is non-empty provided there exists at least one measurable σ_{-i} that satisfies (1). Such σ_{-i} exist whenever F is non-empty and measurable, and more generally whenever F admits a measurable selection.*

Given F as in the previous definition, with non-empty $F_{t_j} \subset A_{t_j}$ for all t_j and $j \neq i$, we will write $\Psi_i^{\mathcal{T}}(t_i, F)$ for the set of beliefs on the finite set $A \times \Theta$, consistent with type t_i's beliefs and certainly that other players are choosing actions consistent with F_{-i}. Thus

$$\Psi_i^{\mathcal{T}}(t_i, F) = \left\{ \psi_i \in \Delta(A_{-i} \times \Theta) \;\middle|\; \begin{array}{l} \psi_i[(\theta, a_{-i})] = \int\limits_{T_{-i}} \sigma_{-i}(t_{-i}, \theta)[a_{-i}] \, \pi(t_i)[(\mathrm{d}t_{-i}, \theta)] \\ \text{for some measurable } \sigma_{-i} : T_{-i} \times \Theta \to \Delta(A_{-i}) \\ \text{such that } \sigma_{-i}(t_{-i}, \theta)[a_{-i}] > 0 \Rightarrow a_{-i} \in F_{t_{-i}} \end{array} \right\}. \tag{4}$$

This is the set of distributions over $A \times \Theta$ that is consistent with t_i's beliefs about $T_{-i} \times \Theta$ and certainty that the play of t_{-i} is consistent with F_{-i}, so

$$BR_i^{\mathcal{T}}(t_i, F) = \left\{ a_i \in A_i \;\middle|\; \begin{array}{c} \text{there exists } \psi_i \in \Psi_i^{\mathcal{T}}(t_i, F) \text{ such that} \\ a_i \in \arg\max\limits_{a_i'} \sum\limits_{A_{-i} \times \Theta} \psi_i[(a_{-i}, \theta)] \, g_i((a_i', a_{-i}), \theta) \end{array} \right\}.$$

4.2.2 Iterative definitions

We now define rationalizability as the result of iterating the BR map. As in the finite case, let $R_0^{\mathcal{T}} = \left((A_i)_{t_i \in T_i}\right)_{i \in I}$, $R_k^{\mathcal{T}} = BR^{\mathcal{T}}\left(R_{k-1}^{\mathcal{T}}\right)$, and $R^{\mathcal{T}} = \cap_{k=1}^{\infty} R_k^{\mathcal{T}}$. The corresponding objects on the universal type space are $R_0^* = \left((A_i)_{t_i^* \in T_i^*}\right)_{i \in I}$, $R_k^* = BR^{\mathcal{T}^*}\left(R_{k-1}^*\right)$, and $R^* = \cap_{k=1}^{\infty} R_k^*$. Let $\Psi_{i,k}^{\mathcal{T}}(t_i) = \Psi_i^{\mathcal{T}}(t_i, R_k^{\mathcal{T}})$, and $\Psi_{i,k}^*(t_i^*) = \Psi_i^{\mathcal{T}^*}(t_i^*, R_k^*)$.

Lemma 1 *If φ is a belief-preserving morphism from (T, π) to (T^*, π^*) and $\varphi_i(t_i) = t_i^*$, then for all k, $R_{i,k}^{\mathcal{T}}(t_i) = R_{i,k}^*(t_i^*)$, $\Psi_i^{\mathcal{T}}\left(t_i, R_{-i,k}^{\mathcal{T}}\right) = \Psi_i^*(\varphi_i(t_i), R_{-i,k}^{\mathcal{T}^*})$, $R_{i,k}^{\mathcal{T}} : T_i \to 2^{A_i}/\emptyset$ is a measurable function, and $\left\{t_i^* \in T_i^* : a_i \in R_{i,k}^*(t_i^*)\right\}$ is closed in the weak topology.*

Proof. The proof is by induction on k. Endow the universal type space with the product topology, where each level of the beliefs is given the weak topology (as in the usual topological construction of the universal type space), and suppose the claim has been shown for all

$k' \leq k$. So suppose that for all i, and $t_i \in T_i$, $R^{\mathcal{T}}_{i,k-1}(t_i) = R^*_{i,k-1}(\varphi(t_i))$ and $\Psi^{\mathcal{T}}_{i,k-1}(t_i) = \Psi^*_{i,k-1}(\varphi(t_i))$, that $R^{\mathcal{T}}_{i,k} : T_i \to 2^{A_i}/\emptyset$ is a measurable function, and that $\left\{t_i^* : a_i \in R^*_{i,k}(t_i^*)\right\} \subset T_i^*$ is closed.

(Part I) The set $\left\{t_i^* : a_i \in R^*_{i,k}(t_i^*)\right\}$ is closed and therefore measurable. To see this, consider a sequence t_i^{*n} that converges to t_i^* and such that $a_i \in R^*_{i,k}(t_i^{*n})$. Then for each t_i^{*n} there is a $\psi_i^{k-1,n} \in \Psi^*_{i,k-1}(t_i^{*n})$ such that $a_i \in \arg\max_{a_i'} \sum_{A_{-i}\times\Theta} g_i(a_i', a_{-i}, \theta)\, \psi_i^{k-1,n}[(a_{-i}, \theta)]$. Moreover, $\psi_i^{k-1,n}[(\theta, a_{-i})] = \int_{T_{-i}} \sigma_{-i}^{*n}(t_{-i}, \theta)[a_{-i}]\, \pi_i(t_i^{*n})[(\mathrm{d}t_{-i}, \theta)]$ for some $\sigma_{-i}^{*n} : T_{-i}^* \times \Theta \to \Delta(A_{-i})$ where $\sigma_{-i}^{*n}(t_{-i}^*, \theta)[a_{-i}] > 0$ implies $a_{-i} \in R^*_{-i,k-1}(t_{-i}^*)$. Let $\nu^{*n} = \nu\left(\sigma_{-i}^{*n}, \pi_i(t_i^{*n})\right)$, and by compactness of $\Delta\left(T_{-i}^* \times \Theta \times A_{-i}\right)$ consider a convergent subsequence of ν^{*n} converging to $\nu^{\infty*}$. Moreover, by compactness we also have a regular version of conditional probabilities, denoted $\nu^{\infty*}[\cdot|(t_{-i}, \theta)] \in \Delta(A_{-i})$, which, by regularity is a measurable function on $T_{-i}^* \times \Theta$. Hence we can define a measurable $\sigma_{-i}^{\infty*} : T_{-i}^* \times \Theta \to \Delta(A_{-i})$ by $\sigma_{-i}^{\infty*}(t_{-i}, \theta)[a_{-i}] = \nu^{\infty*}[a_{-i}|(t_{-i}, \theta)]$. Note that $\nu^{\infty*} = \nu\left(\pi_i(t_i^*), \sigma_{-i}^{\infty}\right)$. Define $\psi_i \in \Delta(A_{-i} \times \Theta)$ by $\psi_i[(a_{-i}, \theta)] \equiv \int_{T_{-i}^*} \nu^{\infty*}[(\mathrm{d}t_{-i}, \theta, a_{-i})]$. Clearly $a_i \in \arg\max_{a_i'} \sum_{A_{-i}\times\Theta} g_i(a_i', a_{-i}, \theta)\, \psi_i[a_{-i}, \theta]$. It remains to show that $\sigma_{-i}^{\infty*}(t_{-i}, \theta)[a_{-i}] > 0 \Rightarrow a_{-i} \in R^*_{-i,k-1}$.

Note first that

$$\nu^{\infty*}\left[\left\{(t_{-i}, \theta, a_{-i}) : a_{-i} \in R^*_{-i,k-1}\right\}\right] = 1. \tag{5}$$

This follows from $\nu^{*n}\left[\left\{\left(t_{-i}^*, \theta, a_{-i}\right) : a_{-i} \in R^*_{-i,k-1}\right\}\right] = 1$ and $\nu^{*n} \to \nu^{\infty*}$.

Equation (5) can be written as $\pi(t_i^*)[N] = 0$, where $N \equiv \left\{(t_{-i}, \theta) : \operatorname{supp}\sigma_{-i}^{\infty}(t_{-i}, \theta) \not\subset R^*_{-i,k-1}\right\}$. So changing $\sigma_{-i}^{*\infty}$ on N has no effect on any expect payoffs or other calculations, and can be done so long as measurability of $\sigma_{-i}^{*\infty}$ continues to be satisfied. Fix $\theta \in \Theta$ for the remainder of the argument. For each (of the finitely many) $B \subset A_{-i}$, let $B^* \equiv \left\{t_{-i} \in T_{-i}^* : R^*_{-i,k-1} = B\right\}$ and $B^\sigma \equiv \left\{t_{-i} \in T_{-i}^* : \operatorname{supp}\sigma_{-i}^{\infty}(t_{-i}, \theta) \subset B\right\}$. Both sets are measurable, hence $B^* - B^{\sigma_i}$ is measurable, and since $\pi(t_i^*)[N]$ also $\pi(t_i^*)[B^* - B^\sigma] = 0$. So redefine $\sigma_{-i}^{*\infty}(t_{-i}, \theta)$ on $B^* - B^\sigma$ to equal any $a_{-i} \in B$. Since $\{a_{-i}\}^\sigma$ is measurable, so is $\{a_{-i}\}^\sigma \cup (B^* - B^\sigma)$, so after this redefinition $\sigma_{-i}^{*\infty}$ is still measurable and $B^* - B^\sigma$ is empty. Doing this process for all $B \subset A_{-i}$ we obtain a measurable $\sigma_{-i}^{*\infty}$ such that $\sigma_{-i}^{*\infty}(t_{-i}, \theta) \in R^*_{-i,k-1}$ for every (not only a.e.) t_{-i}.

(Part II) Assume $a_i \in R^{\mathcal{T}}_{i,k}(t_i)$. Since $\Psi^{\mathcal{T}}_{i,k-1}(t_i) = \Psi^*_{i,k-1}(\varphi(t_i))$ it is immediate that $R^{\mathcal{T}}_{i,k}(t_i) = R^*_{i,k}(\varphi_i(t_i))$.

(Part III) By (part I), (part II) and the measurability of φ_i we have that $R^{\mathcal{T}}_{i,k} : T_i \to 2^{A_i}/\emptyset$ is measurable.

(Part IVa) We now argue that $\Psi^*_{i,k}(t_i^*) \subset \Psi^{\mathcal{T}}_{i,k}(t_i)$.

$$\Psi^{*}_{i,k}\left(t^{*}_{i}\right)=\left\{\psi^{*}_{i}\in\Delta\left(A_{-i}\times\Theta\right)\,\middle|\,\begin{array}{l}\psi^{*}_{i}\left[\left(\theta,a_{-i}\right)\right]=\int_{T^{*}_{-i}}\sigma^{*}_{-i}\left(t^{*}_{-i},\theta\right)\left[a_{-i}\right]\pi^{*}_{i}\left(t^{*}_{i}\right)\left[\left(\mathrm{d}t^{*}_{-i},\theta\right)\right]\\ \text{for some measurable }\sigma^{*}_{-i}:T^{*}_{-i}\times\Theta\to\Delta\left(A_{-i}\right)\\ \text{such that }\sigma^{*}_{-i}\left(t^{*}_{-i},\theta\right)\left[a_{-i}\right]>0\Rightarrow a_{-i}\in R^{*}_{-i,k}\end{array}\right\}$$

Fix ψ^{*}_{i} and the σ^{*}_{-i} in the above expression, and define $\sigma_{-i}:T_{-i}\times\Theta\to\Delta\left(A_{-i}\right)$ by $\sigma_{-i}\left(t_{-i},\theta\right)=\sigma^{*}_{-i}\left(\varphi_{-i}\left(t_{-i}\right),\theta\right)$. Since $R^{\mathcal{T}}_{-i,k}\left(t_{-i}\right)=R^{*}_{-i,k}\left(\varphi_{-i}\left(t_{-i}\right)\right)$ and $\sigma^{*}_{-i}\left(t^{*}_{-i},\theta\right)\left[a_{-i}\right]>0\Rightarrow a_{-i}\in R^{*}_{-i,k}$ we have $\sigma_{-i}\left(t_{-i},\theta\right)\left[a_{-i}\right]>0\Rightarrow a_{-i}\in R^{\mathcal{T}}_{-i,k}$. So

$$\psi_{i}\left[\left(\theta,a_{-i}\right)\right]=\int\limits_{T_{-i}}\sigma_{-i}\left(t_{-i},\theta\right)\left[a_{-i}\right]\pi_{i}\left(t_{i}\right)\left[\left(\mathrm{d}t_{-i},\theta\right)\right]$$

is in $\Psi^{\mathcal{T}}_{i}\left(t_{i},R^{\mathcal{T}}_{-i,k}\right)$. ¿From the morphism we have that

$$\begin{aligned}\psi^{*}_{i}\left[\left(\theta,a_{-i}\right)\right]&=\int\limits_{T^{*}_{-i}}\sigma^{*}_{-i}\left(\varphi_{-i}\left(t_{-i}\right),\theta\right)\left[a_{-i}\right]\pi^{*}_{i}\left(\varphi\left(t_{i}\right)\right)\left[\left(\mathrm{d}t^{*}_{-i},\theta\right)\right]\\&=\int\limits_{T_{-i}}\sigma_{-i}\left(t_{-i},\theta\right)\left[a_{-i}\right]\pi_{i}\left(t_{i}\right)\left[\left(\mathrm{d}t_{-i},\theta\right)\right].\end{aligned}$$

Thus

$$\Psi^{*}_{i,k}\left(t^{*}_{i}\right)\subset\Psi^{\mathcal{T}}_{i,k}\left(t_{i}\right).$$

(Part IVb) To prove the converse, suppose $\psi\in\Psi^{\mathcal{T}}_{i}\left(t_{i},R^{\mathcal{T}}_{-i,k}\right)$ and let σ_{-i} be the associated conjecture so $\psi_{i}\left[\left(\theta,a_{-i}\right)\right]=\int_{T_{-i}}\sigma_{-i}\left(t_{-i},\theta\right)\left[a_{-i}\right]\pi_{i}\left(t_{i}\right)\left[\left(\mathrm{d}t_{-i},\theta\right)\right]$. We will define $\sigma^{*}_{-i}:T^{*}_{-i}\times\theta\to\Delta\left(A_{-i}\right)$ as follows.

First, for every $B_{-i}\subset A_{-i}$ let $T^{B_{-i}}_{-i}=\left\{t_{-i}\in T_{-i}:B_{-i}=R^{\mathcal{T}}_{-i,k-1}\left(t_{-i}\right)\right\}$. Note that by the induction hypothesis, for all $t^{*}_{-i}\in\varphi_{-i}\left(T^{B_{-i}}_{-i}\right)$ we have $R^{\mathcal{T}}_{-i,k-1}\left(t^{*}_{-i}\right)=B_{-i}$. Hence $\varphi_{-i}\left(T^{B_{-i}}_{-i}\right)$ is measurable (see part I). Moreover, by induction we have that if $R^{\mathcal{T}}_{-i,k-1}\left(t_{-i}\right)\neq R^{\mathcal{T}}_{-i,k-1}\left(t'_{-i}\right)$ then $\varphi_{-i}\left(t_{-i}\right)\neq\varphi_{-i}\left(t'_{-i}\right)$. Hence $\left\{\varphi_{-i}\left(T^{B_{-i}}_{-i}\right)\right\}_{B_{-i}\subset A_{-i}}$ is a finite measurable partition of T^{*}_{-i}. Finally, for any B_{-i} fix some $\bar{a}^{B_{-i}}_{-i}\in B_{-i}$.

The idea used to define $\sigma^{*}_{-i}\left(t^{*}_{-i},\theta\right)\left[\cdot\right]\in\Delta\left(A_{-i}\right)$ is as follows. We map $\sigma_{-i}\left(t_{-i},\cdot\right)$ into $\sigma^{*}_{-i}\left(t^{*}_{-i},\cdot\right)$ by taking all t_{-i} for whom B_{-i} is $k-1$ rationalizable, denoted $T^{B_{-i}}_{-i}$, taking the conditional average of $\sigma_{-i}\left(t_{-i},\cdot\right)$ over those t_{-i}, and assigning that average conjecture to those t^{*}_{-i} who have that same $k-1$ rationalizable set. Moreover, those t^{*}_{-i} are the image of $T^{B_{-i}}_{-i}$, and these images, $\varphi_{-i}\left(T^{B_{-i}}_{-i}\right)$, partition T^{*}_{-i}. So we can combine all those averages to get a strategy for all $t^{*}_{-i}\in T^{*}_{-i}$. There is a slight issue for the case where the conditional isn't well defined because the conditioning event, $T_{-i}^{B_{-i}}$, has probability zero. In that case the strategy is really irrelevant, but as we require it to be measurable and to map into the $k-1$ rationalizable set, we add that restriction by having the strategy assign probability 1 to some $k-1$ rationalizable action for all $t_{-i}^* \in \varphi_{-i}\left(T_{-i}^{B_{-i}}\right)$ whenever $\pi_i\left(t_i\right)\left[T_{-i}^{B_{-i}}\right] = 0$.

We now formalize this verbal description.

$$\sigma_{-i}^*\left(t_{-i}^*,\theta\right)\left[a_{-i}\right] = \begin{cases} \frac{\int_{T_{-i}^{B_{-i}}} \sigma_{-i}(t_{-i},\theta)[a_{-i}]\pi(t_i)[(\mathrm{d}t_{-i},\theta)]}{\pi(t_i)\left[T_{-i}^{B_{-i}}\right]} \text{ if } t_{-i}^* \in \varphi\left(T_{-i}^{B_{-i}}\right) \text{ and } \pi\left(t_i\right)\left[T_{-i}^{B_{-i}}\right] > 0 \\ 1 \text{ if } t_{-i}^* \in \varphi\left(T_{-i}^{B_{-i}}\right), \pi\left(t_i\right)\left[T_{-i}^{B_{-i}}\right] = 0 \text{ and } a_{-i} = \bar{a}_{-i} \\ 0 \text{ if } t_{-i}^* \in \varphi\left(T_{-i}^{B_{-i}}\right), \pi\left(t_i\right)\left[T_{-i}^{B_{-i}}\right] = 0 \text{ and } a_{-i} \neq \bar{a}_{-i} \end{cases}$$

This is measurable because it is constant on each of the finitely many measurable cell$s$ of $\left\{\varphi\left(T_{-i}^{B_{-i}}\right)\right\}_{B_{-i}\subset A_{-i}}$. Moreover, $\sigma_{-i}^*\left(t_{-i}^*,\theta\right)\left[a_{-i}\right] > 0 \Rightarrow a_{-i} \in R_{-i,k-1}^{\mathcal{T}^*}$. So this σ_{-i}^* can be used to define $\psi_i^* \in \Psi_i^*\left(t_i^*, R_{-i,k}^*\right)$ by $\psi_i^*\left[\theta, a_{-i}\right] = \int_{T_{-i}^*} \sigma_{-i}^*\left(t_{-i}^*,\theta\right)\left[a_{-i}\right]\pi_i^*\left(t_i^*\right)\left[\left(\mathrm{d}t_{-i}^*,\theta\right)\right]$.

Now

$$\begin{aligned} \psi_i^*\left[\theta, a_{-i}\right] &= \int\limits_{T_{-i}^*} \sigma_{-i}^*\left(\varphi_{-i}\left(t_{-i}\right),\theta\right)\left[a_{-i}\right]\pi_i^*\left(\varphi_i\left(t_i\right)\right)\left[\left(\mathrm{d}\left(\varphi_{-i}\left(t_{-i}\right)\right),\theta\right)\right] = \\ &= \sum_{B_{-i}} \left(\frac{\int_{T_{-i}^{B_{-i}}} \sigma_{-i}\left(t_{-i},\theta\right)\left[a_{-i}\right]\pi_i\left(t_i\right)\left[\left(\mathrm{d}t_{-i},\theta\right)\right]}{\pi\left(t_i\right)\left[T_{-i}^{B_{-i}}\right]}\right)\pi_i^*\left(\varphi\left(t_i\right)\right)\left[\varphi_{-i}\left(T_{-i}^{B_{-i}}\right)\right] \\ &= \int\limits_{T_{-i}} \sigma_{-i}\left(t_{-i},\theta\right)\left[a_{-i}\right]\pi_i\left(t_i\right)\left[\left(\mathrm{d}t_{-i},\theta\right)\right] = \psi_i\left[\theta, a_{-i}\right], \end{aligned}$$

where the first equality is by definition, the second by substitution and changing the integration of a finite valued "step function" to a sum, and the third using $\pi_i^*\left(\varphi\left(t_i\right)\right)\left[\varphi_{-i}\left(T_{-i}^{B_{-i}}\right)\right] = \pi\left(t_i\right)\left[T_{-i}^{B_{-i}}\right]$. Thus for any $\psi \in \Psi_i^{\mathcal{T}}\left(t_i, R_{-i,k}^{\mathcal{T}}\right)$ we have found $\psi^* \subset \Psi_i^*\left(t_i^*, R_{-i,k}^*\right)$. ■

Corollary 1 *$R_i^{\mathcal{T}}\left(t_i\right) = R_i^{\mathcal{T}}\left(t_i'\right)$ if $\hat{\pi}_i^*(t_i) = \hat{\pi}_i^*(t_i')$.*

4.2.3 Fixed-point definitions

Modulo the requirement that σ_{-i} be measurable, and replacing summations with integrals, the definition of best reply sets is in the finite case. The properties mentioned there also hold, although the argument is slightly different.

Lemma 2 *1. If $S_c^{\mathcal{T}}$ for all c in some index set C are best-reply sets then $\cup_c S_c^{\mathcal{T}}$ is a best-reply set.*

2. The union of all best-reply sets is the largest fixed point of $\mathrm{BR}^{\mathcal{T}}$.

To see property 2 denote the union of all best-reply sets as $\mathcal{S}$ and observe that if $a_i \in BR_i^{\mathcal{T}}(t_i, \mathcal{S}_{-i})$, then adding a_i to $\mathcal{S}_i(t_i)$ will continue to constitute a best-reply set.

Definition 4 $R_F^{\mathcal{T}} = \left(\left(R_{i,\mathcal{F}}^{\mathcal{T}}(t_i)\right)_{t_i \in T_i}\right)_{i \in I} \subset \left(A^{T_i}\right)_{i \in I}$ *is the largest fixed point of* $\mathrm{BR}^{\mathcal{T}}$.

In general, the largest fixed point need not be coincide with the iterative definition given above, as reducing the set to the largest fixed point may require transfinite induction; see Lipman (1994). However, because payoffs depend only on distributions over the finite sets of actions and states of nature, we can show that the fixed point definition is well posed and coincides with the iterative definition.

Proposition 4 $R_{\mathcal{F}}^{\mathcal{T}}$ *equals* $R^{\mathcal{T}}$.

Proof. It is sufficient to prove that $R^{\mathcal{T}}$ is a best-reply set. That nothing larger can be a best-reply set is immediate. For every $a \in R^{\mathcal{T}}$ we have that for every k there is a conjecture $\nu_k \in \Delta\left(R_{-i,k}^{\mathcal{T}} \times \Theta\right)$ against which a is a best reply. Since all the $R_{i,k}^{\mathcal{T}}$ are closed, and u is continuous, we have that there exists a limit to a convergent subsequence of ν_k, such that a is best reply against $\nu \in R_{-i}^{\mathcal{T}}$. ■

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

References

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

[2] —— and A. Brandenburger (1995). "Epistemic Conditions for Nash Equilibrium," *Econometrica* **63**, 1161–1180.

[3] Battigalli, P. and M. Siniscalchi (1999). "Hierarchies of Conditional Beliefs and Interactive Epistemology in Dynamic Games," *Journal of Economic Theory* **88**, 188-230

[4] —— (2003). "Rationalization and Incomplete Information," *Advances in Theoretical Economics* **3**, Article 3. http://www.bepress.com/bejte/advances/vol3/iss1/art3/.

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

[6] Bergemann, D. and S. Morris (2005). "Robust Mechanism Design," *Econometrica* **73**, 1521-1534.

[7] Billingsley, P. (1995). *Probability and Measure*, 3rd ed. John Wiley and Sons, New York, NY, USA.

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

[9] —— (1993). "Hierarchies of Beliefs and Common Knowledge," *Journal of Economic Theory* **59**, 189-198.

[10] Dekel, E., D. Fudenberg and S. Morris (2006). "Topologies on Types," forthcoming in *Theoretical Economics.*

[11] Ely, J. and M. Peski (2005). "Hierarchies of Belief and Interim Rationalizability," forthcoming in *Theoretical Economics.*

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

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

[14] 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.

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

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

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

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

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

[20] Pearce, D. (1984). "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica* **52**, 1029-1051.

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

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

[23] Yildiz, M. (2006). "Generic Uniqueness of Rationalizable Outcomes."