Theoretical Economics 2 (2007), 15–40 1555-7561/20070015

Interim correlated rationalizability

Eddie Dekel
Department of Economics, Northwestern University, and School of Economics, Tel Aviv University

Drew Fudenberg
Department of Economics, Harvard University

Stephen Morris
Department of Economics, Princeton University

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 certainty of rationality in the universal type space.

Keywords. Rationalizability, incomplete information, common certainty, common knowledge, universal type space.

JEL classification. C70, C72.

1. Introduction

Harsanyi (1967–68) 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.

Eddie Dekel: dekel@northwestern.edu
Drew Fudenberg: fudenberg@husc.harvard.edu
Stephen Morris: smorris@Princeton.edu

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 for helpful comments, and are particularly grateful to Aviad Heifetz and Muhamet Yildiz for helpful discussions, Paolo Battigalli, Bart Lipman, Marcin Pęski, and the referees 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.

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

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 show also 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 certainty of rationality.[3] Third, we extend a result of Brandenburger and Dekel (1987). They show 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 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 probability distribution over 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. We call such probability distributions "forecasts." In the case of interim independent rationalizability, the allowed forecasts for 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 interim correlated rationalizability, the allowed forecasts are generated by combining t's beliefs over types and Θ with any, perhaps correlated, conditional conjectures about which (surviving) actions

[2] We use the concept of interim correlated rationalizability in our study of topologies on the universal type space (Dekel et al. 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.

[3] We follow a recent convention in the epistemic foundations of game theory literature of using "certainty" to mean "belief with probability one." An earlier convention—that we followed in Dekel et al. (2006) and other earlier work—uses "knowledge" instead to refer belief with probability one. But the convention in philosophy and other closely related disciplines has been to reserve "knowledge" for true belief. The truth of players' beliefs does not play a role in our analysis, so in this paper we switch to the newer convention.

are played at a given type profile and payoff-relevant state. In this latter definition, a type's forecast can allow for correlation among the payoff-relevant state, other players' types, and other players' actions.[4]

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 shed little light on either the motivation for the definition 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 certainty 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 Section 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
U	$1,0$	$0,0$
D	$\frac{3}{5},0$	$\frac{3}{5},0$

θ':	L	R
U	$0,0$	$1,0$
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 certainty.[5] Clearly, either action is rational for player 2, as she is indifferent between both 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

[4]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 forecast of the opponent's actions and the payoff-relevant state, conditional on the opponent's type.

[5]Formally, this means that the event that each player assigns equal probability to the states is common certainty, as defined in Section 3.4.

be θ' and player 2 will choose R. This makes U optimal for player 1. As we will see in Section 3.4, this means that U is consistent with common certainty of rationality.

Action U is also consistent with interim correlated rationalizability. To illustrate this we consider two type spaces that capture the same assumptions as above about the players' higher-order beliefs. In type space $\mathscr{T}$, each player $i = 1,2$ has two possible types, $T_i = \{t_i', t_i''\}$ and beliefs are generated by the following common prior over $T_1 \times T_2 \times \{\theta, \theta'\}$:

$$\theta: \quad \begin{array}{c|c|c|} & t_2' & t_2'' \\ \hline t_1' & \frac{1}{6} & \frac{1}{12} \\ \hline t_1'' & \frac{1}{12} & \frac{1}{6} \\ \hline \end{array} \qquad \theta': \quad \begin{array}{c|c|c|} & t_2' & t_2'' \\ \hline t_1' & \frac{1}{12} & \frac{1}{6} \\ \hline t_1'' & \frac{1}{6} & \frac{1}{12} \\ \hline \end{array}$$

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

$$\theta: \quad \begin{array}{c|c|} & \hat{t}_2 \\ \hline \hat{t}_1 & \frac{1}{2} \\ \hline \end{array} \qquad \theta': \quad \begin{array}{c|c|} & \hat{t}_2 \\ \hline \hat{t}_1 & \frac{1}{2} \\ \hline \end{array}$$

Notice that in both type spaces, for *every* type of both players, there is common certainty that each player assigns probability $\frac{1}{2}$ to the true state being θ. The types in $\mathscr{T}$ 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 conjectures about their opponents and this is potentially important depending on the choice of solution concept. Redundant types can serve as a correlating device, and so these types are not truly "redundant" unless the addition of correlating devices has no effect.

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 forecast 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 is 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 forecast over (t_j, θ, a_j) must treat the choice of player j's action as independent of θ, conditional on his type. With this solution concept, action U is not interim independent rationalizable for type $\hat{t}_1$: there is no conditionally independent forecast over actions, states, and types 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' conjectures that type t_2' will play action L and type t_2'' will play action R, then he attaches probability $\frac{1}{3}$ to each of action-state profiles (L, θ) and (R, θ'). This is enough to make action U a best response. Thus both U and D are interim independent rationalizable for types t_1' and t_1''. $\Diamond$

The example 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 forecasts over other players' actions, their types, 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.

Three papers 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-certainty restrictions on players' forecasts. They show that there is an equivalence between actions surviving an iterative procedure capturing common certainty of Δ and the set of actions that might be played in equilibrium on any type space where Δ is common certainty. 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.[6]

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 forecasts that we use in defining our solution concept (we discuss this further in Section 3.2). Thus our approach is the non-common prior analogue of Forges' universal Bayesian approach.

A recent paper by Ely and Pęski (2006) also notes that the set of interim independent rationalizable outcomes in two-player games depends on more than just the hierarchy of beliefs over the payoff-relevant states of Nature. In response, they provide an extended notion of hierarchies of beliefs for 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 players, I, and for each player $i \in I$ a finite set of actions A_i and a payoff function g_i, where $g_i : A \times \Theta \to [0,1]$ and $A = (A_i)_{i \in I}$.[7] For the first part of the paper, we restrict attention to a finite type space $\mathscr{T} = (T_i, \pi_i)_{i \in I}$, 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 $T_{-i} \times \Theta$.[8] We later relax the assumption that $\mathscr{T}$ 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.

[6] The common-certainty restrictions Δ in Battigalli and Siniscalchi (2003) are assumed to restrict only first-order beliefs about Θ, not higher-order beliefs. Thus while our exercise is conceptually a special case of Battigalli and Siniscalchi (2003), a slightly extended class of restrictions Δ would be required to formally incorporate it.

[7] Henceforth we use analogous notation, e.g., $Q = (Q_c)_{c \in C}$ for $\{Q_c : c \in C\}$. 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 sets are written as usual as $q_c \in Q_c$, $q \in Q$, and $q_{-i} \in Q_{-i}$.

[8] Throughout the paper, every finite set is given the obvious 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. Thus we call π_i player i's belief.

The main solution concept that we study is that of interim correlated rationalizability, or ICR.[9] As with correlated rationalizability in complete-information games, ICR is defined by an iterative deletion procedure. At each round, an action survives for a given type only if it is a best response to a forecast 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 (2) is consistent with that type's beliefs about $T_{-i}\times\Theta$. Formally, we have $R^{\mathcal{T}}_{i,0}(t_i)=A_i$,

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

and $R^{\mathcal{T}}_i(t_i)=\cap_{k=1}^{\infty}R^{\mathcal{T}}_{i,k}(t_i)$.

To better explain ICR we compare it to a related solution concept, interim independent rationalizability, or IIR. The latter concept imposes the additional restriction that type t_i's forecast supporting an action corresponds to independent conjectures, i.e., that the conjecture about actions conditional on type and state is a product measure that does not depend on the state.

Let $IIR^{\mathcal{T}}_{i,0}(t_i)=A_i$,

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

and $IIR^{\mathcal{T}}_i(t_i)=\cap_{k=1}^{\infty}IIR^{\mathcal{T}}_{i,k}(t_i)$.

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, which we mention below but do not define formally, could allow for correlation among players' actions but not with

[9]We discuss the sense in which these concepts are interim and how they relate to an ex ante concept in the concluding remarks.

Nature, by specifying $\sigma_{-i}: T_{-i} \to \Delta(A_{-i})$ instead of $(\sigma_j)_{j\neq i}$ in (3) above in the definition of IIR.

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

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

For each $k = 2, 3, \ldots$, let

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

Finally, let

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

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

PROOF. We establish by induction for each k that if $\widehat{\pi}_i^k(t_i) = \widehat{\pi}_i^k(t_i')$ 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

$$(1)\ \nu[(t_{-i},\theta,a_{-i})] > 0 \Rightarrow a_{-i} \in (R_{j,k-1}^{\mathcal{T}}(t_j))_{j\neq i}$$
$$(2)\ a_i \in \arg\max_{a_i'} \sum_{t_{-i},\theta,a_{-i}} g_i((a_i',a_{-i}),\theta)\nu[(t_{-i},\theta,a_{-i})]$$
$$(3)\ \sum_{a_{-i}} \nu[(t_{-i},\theta,a_{-i})] = \pi_i(t_i)[(t_{-i},\theta)].$$

We now construct $\nu' \in \Delta(T_{-i}'\times\Theta\times A_{-i})$ such that the above three conditions hold when ν' and $\pi_i(t_i')$ replace ν and $\pi_i(t_i)$, respectively. Let $D_{-i}^{k-1} = \{(\widehat{\pi}_j^{k-1}(t_{-i}))_{j\neq i} : t_{-i} \in T_{-i}\}$. For $\widetilde{\pi}_{-i}^k \in D_{-i}^{k-1}$, let

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

Then for $(\widetilde{\pi}_{-i}^k,\theta)$ such that $\gamma(\widetilde{\pi}_{-i}^{k-1},\theta) > 0$, set

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

and for all other $(\widetilde{\pi}_{-i}^{k-1}, \theta)$ set

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

Next, let

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

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

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

Hence we obtain the following.

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

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

Now we claim

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

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

The intuition for the proposition is as follows. 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 the forecast then depends only on first-order beliefs $\widehat{\pi}_i^1$ about Θ. Second-level rationalizability depends on beliefs about Θ and the opponents' first-level rationalizable sets; these in turn depend only on the opponents' first-order beliefs, so second-level rationalizability is determined by second-order beliefs, $\widehat{\pi}_i^2$, and so on.

In the course of the proof we demonstrated the corollary below, which gives an equivalent definition of $R^{\mathcal{T}}_{i,k+1}(t_i)$. It states that t_i's forecast ν can be decomposed into conjectures σ_i about opponents' strategies, and the beliefs π_i.

COROLLARY 1.

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

Note that i's conjecture about other players' actions, σ_i, allows for j's action to be correlated with other players' actions, the state, and other players' types. Contrasting this with the definition of IIR makes clear where ICR allows additional correlation.

3.2 *Discussion*

The correlation allowed by ICR can have surprising consequences, as in the next example.

EXAMPLE 2. There are two states of Nature θ and θ', and it is common certainty that each player assigns probability $\frac{1}{2}$ to each state. Thus in the universal type space each player $i = 1,2$ has a single type t_i^*. 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. To see this, note that N is a best response to the forecast that assigns probabilities $\frac{1}{2}$ to (t_2^*,θ,N) and $\frac{1}{2}$ to (t_2^*,θ',N) (which implies that the opponent always plays N), that B is a best reply for player 1 to the forecast $\frac{1}{2}$ to (t_2^*,θ,B) and $\frac{1}{2}$ to (t_2^*,θ',N), and symmetrically that B is a best reply for player 2 to the forecast $\frac{1}{2}$ to (t_1^*,θ,N) and $\frac{1}{2}$ to (t_2^*,θ',B). Thus the ICR set for each player is $\{B,N\}$. Note that the forecast that supports B for player i supposes that the opposing player j bets exactly when this is good for i and bad for j.

Thus each player expects there to be costly speculative trade (and indeed using the epistemic set-up of the next section, there is common certainty of trade with probability bounded above zero) 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.[10] $\Diamond$

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) show that correlated rationalizability (allowing players to have correlated conjectures over others' actions) corresponds to common certainty 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, p. 16):

> 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?

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.[11]

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.[12] A difficulty with

[10] Although there is a common prior over Θ, this observation is not inconsistent with no-trade theorems because there is not common certainty of the conditional probability of trade in each state θ. This lack of common certainty 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 et al. (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.

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

[12] Ely and Pęski (2006) 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

such hybrid notions is that the resulting solution concept is 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 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's 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 forecasts about players' actions but not 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.[13] ◊

A recent paper of Brandenburger and Friedenberg (2006) suggests a solution concept intermediate between correlated rationalizability and independent rationalizability in complete-information games. They require that players hold conditionally independent conjectures about their opponents' play, contingent on their beliefs and higher-order beliefs about players' actions, and also that a player's conjecture about another player's actions does not change if he learns a third player's beliefs and higher-order beliefs about players' actions. They show that most, but not all, correlated-rationalizable actions satisfy common certainty of rationality and these restrictions. Intuitively, higher-order uncertainty about players' actions introduces intrinsic correlation into the game environment. One could presumably extend their solution concept to incomplete information settings to obtain yet another solution concept intermediate between interim correlated rationalizability and interim independent rationalizability.

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* As one might expect from earlier work, 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

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.

[13]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 is certain that player 3 is certain of θ. Unlike the example in the text, this version does not reproduce the entire set of ICR actions.

$U^{\mathcal{T}}_{i,0}(t_i) = A_i,$

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

and

$$U^{\mathcal{T}}_i(t_i) = \cap_{k=1}^{\infty} U^{\mathcal{T}}_{i,k}(t_i).$$

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

3.3.2 *Best-reply sets* Similarly, there is an obviously equivalent best-reply set (Pearce 1984) definition of ICR. Let $S^{\mathcal{T}}_i : T_i \to 2^{A_i} \setminus \varnothing$ be a specification of possible actions for each type, and $S^{\mathcal{T}} = (S^{\mathcal{T}}_i)_{i\in I}$.[14]

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

(1) $\sigma_{-i}(t_{-i},\theta)[a_{-i}] > 0 \Rightarrow a_{-i} \in S^{\mathcal{T}}_{-i}(t_{-i})$

(2) $a_i \in \arg\max_{a'_i} \displaystyle\sum_{t_{-i},a_{-i},\theta} \pi_i(t_i)[(t_{-i},\theta)]\sigma_{-i}(t_{-i},\theta)[a_{-i}] g_i((a'_i,a_{-i}),\theta)$.

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

(ii) The union of all best-reply sets is equal to $((R^{\mathcal{T}}_i(t_i))_{t_i\in T_i})_{i\in I}$.

Property (i) is immediate. That the union includes $R^{\mathcal{T}}_i$ follows from the observation that $((R^{\mathcal{T}}_i(t_i))_{t_i\in T_i})_{i\in I}$ 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 $R^{\mathcal{T}}_i$. 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.

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

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

$$BR_i^{\mathscr{T}}(t_i, F) = \left\{ a_i \in A_i : \begin{array}{l} \text{there exists } \sigma_{-i} : T_{-i} \times \Theta \to \Delta(A_{-i}) \text{ such that} \\ (1)\ \sigma_{-i}(t_{-i}, \theta)[a_{-i}] > 0 \Rightarrow a_{-i} \in F_{t_{-i}} \\ (2)\ a_i \in \arg\max\limits_{a_i'} \sum\limits_{t_{-i},\theta,a_{-i}} g_i((a_i', a_{-i}), \theta)\sigma_{-i}(t_{-i},\theta)[a_{-i}]\pi_i(t_i)[(t_{-i},\theta)] \end{array} \right\}.$$

Next we define[15]

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

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

This follows from the fact that any fixed point is a best-reply set and the previous claim regarding best-reply sets.

3.4 *Epistemic foundations*

To better understand the two solution concepts, ICR and IIR, we relate them to common certainty of rationality. In order to do this, we introduce a richer language—an "epistemic model"—to model the certainty of the players. We are able to provide an epistemic foundation for the solution concepts in the spirit of the existing epistemic-foundations literature.[16] We note that—at least in our epistemic formulation—additional common-certainty 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 $\mathscr{T} = (T_i, \pi_i)_{i\in I}$. We sometimes refer to this object as a "standard" type space and to elements of T_i as standard types. As discussed, we view this type space as exogenously given, for example describing the perspective of the modeler of the environment, 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 certainty. That is, we embed the standard type space in an arbitrary larger space, the epistemic space—which can be any extension to a more complete description of the players' perceptions of the world, specifying at least their actions at any state—and we assume that the (original) type space is common certainty in this epistemic space. Then we ask what can we say about play in the game defined by the original type space; i.e., what solution concept—

[15] 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.

[16] Aumann (1987), Brandenburger and Dekel (1987), Tan and Ribeiro da Costa Werlang (1988), Aumann and Brandenburger (1995).

defined on games with the original type space—is characterized by common certainty of rationality.[17]

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, $\mathrm{a}_i : E_i \to A_i$
3. i's "standard type," $\tau_i : E_i \to T_i$.

Thus an epistemic model consists of $(E_i,\phi_i,\mathrm{a}_i,\tau_i)_{i\in I}$; 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",

$$Rat_i = \{((e_i,e'_{-i}),\theta') : \mathrm{a}_i(e_i)\in \arg\max_{a_i}\sum_{e_{-i},\theta} g_i((a_i,\mathrm{a}_{-i}(e_{-i})),\theta)\phi_i(e_i)[(e_{-i},\theta)]\},$$

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:

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

$$W = \bigcap_i W_i.$$

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

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

the set of states where everyone is certain of the event H is

$$C_*(H) = \bigcap_i C_i(H),$$

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

$$CC(H) = \bigcap_{n=0}^{\infty} (C_*)^n(H),$$

where $(C_*)^0(H)=H$.

[17] In general it would be reasonable to allow for a larger space of states of Nature $\Theta'\supset\Theta$, where $\Theta'\setminus\Theta$ are payoff irrelevant, as Θ is also a partial, "small worlds," description of the payoff-relevant parameters. It is easy to see that allowing such an enlargement would not affect our results. The same remark holds for adding the complete universe of "dummy" players whose actions do not affect the payoffs of the "small worlds" set I in the game studied. As noted, adding payoff-irrelevant states, dummy players, or just alternative epistemic states as below, does not affect the set of ICR actions but changes which actions are IIR.

PROPOSITION 2. *Interim Correlated Rationalizability characterizes common certainty of rationality and of the standard type space. That is,*

(i) in any epistemic model, if $((e_i, e_{-i}), \theta) \in CC(Rat \cap W)$*, then* $\mathrm{a}_i(e_i) \in R_i^{\mathcal{T}}(\tau_i(e_i))$

(ii) there is an epistemic model such that if $a_i \in R_i^{\mathcal{T}}(t_i)$*, then there is a state* $((e_i, e_{-i}), \theta)$ *such that* $((e_i, e_{-i}), \theta) \in CC(Rat \cap W)$*,* $\tau_i(e_i) = t_i$*, and* $\mathrm{a}_i(e_i) = a_i$*.*

PROOF. (i) Suppose $((e_i^*, e_{-i}^*), \theta^*) \in CC(Rat)$. Let E_j^* be the set of epistemic types of player j where j is certain of $CC(Rat)$. Let

$$S_i(t_i) = \{a_i : \text{ for some } e_i \in E_i^*,\ \mathrm{a}_i(e_i) = a_i \text{ and } \tau_i(e_i) = t_i\}.$$

Observe that, by construction,

$$\mathrm{a}_i(e_i^*) \in S_i(\tau_i(e_i^*)).$$

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

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

Again by construction,

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

Common certainty of W ensures that

$$\sum_{a_{-i}} \lambda_i^{a_i, t_i}[(t_{-i}, \theta, a_{-i})] = \pi_i(t_i)[(t_{-i}, \theta)] \text{ for all } t_{-i}, \theta.$$

Thus an inductive argument ensures that $S_i(t_i) \subseteq R_{i,k}^{\mathcal{T}}(t_i)$ for all k and thus $S_i(t_i) \subseteq R_i^{\mathcal{T}}(t_i)$. So

$$\mathrm{a}_i(e_i^*) \in S_i(\tau_i(e_i^*)) \subseteq R_i^{\mathcal{T}}(\tau_i(e_i^*)).$$

(ii) We construct an epistemic type space. Let $E_i = \{(t_i, a_i) : a_i \in R_i^{\mathcal{T}}(t_i)\}$. Let

$$\begin{aligned} \mathrm{a}_i(e_i) &= \mathrm{a}_i((t_i, a_i)) = a_i \\ \tau_i(e_i) &= \tau_i((t_i, a_i)) = t_i. \end{aligned}$$

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

$$\begin{aligned} &(1)\ \lambda_i^{a_i, t_i}[(t_{-i}, \theta, a_{-i})] > 0 \Rightarrow a_{-i} \in (R_j^{\mathcal{T}}(t_j))_{j \neq i} \\ &(2)\ a_i \in \arg\max_{a_i'} \sum_{t_{-i}, \theta, a_{-i}} \lambda_i^{a_i, t_i}[(t_{-i}, \theta, a_{-i})] g_i((a_i', a_{-i}), \theta) \\ &(3)\ \sum_{a_{-i}} \lambda_i^{a_i, t_i}[(t_{-i}, \theta, a_{-i})] = \pi_i(t_i)[(t_{-i}, \theta)] \text{ for all } t_{-i}, \theta. \end{aligned}$$

Let

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

By construction, $Rat_i = W_i = \Omega$ for all i and thus $CC(Rat \cap W) = \Omega$. Now, also by construction, for any $a_i \in R_i^{\mathcal{T}}(t_i)$, there is an epistemic type $e_i = (t_i, a_i)$ with $((e_i, e_{-i}), \theta) \in CC(Rat \cap W) = \Omega$, $\tau_i(e_i) = t_i$, and $\mathrm{a}_i(e_i) = a_i$. □

REMARK 1. To see why B is consistent with common certainty of rationality in Example 2 note that each player i can believe that two epistemic types of the opposing player can correspond to the same standard type but take different actions, and that the epistemic types are correlated with θ.

REMARK 2. A standard reinterpretation of the result is that if we start with the standard type space $\mathcal{T} = (T_i, \pi_i)_{i\in I}$, we can construct a larger type space $\mathcal{T}' = (T_i', \pi_i')_{i\in I}$ and belief preserving morphisms $\varphi_i : T_i \to T_i'$ from the original type space to the larger type space, and a Nash equilibrium on that larger type space, such that for each type t_i in the original type space and each interim-correlated-rationalizable action for that type, there is a corresponding type $t_i' = \varphi_i(t_i)$ 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 certainty of rationality, every player chooses 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 assumes a singleton type space (i.e. a complete information game) that is embedded in an epistemic space, and shows 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. The approach of the referee corresponds to focusing on the actions that are played on the enlarged game. In two player games, this yields exactly the Nash equilibria. We follow Aumann in studying the implications of common-certainty 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 certainty.

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

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

Let $Y = \bigcap_i Y_i$.

PROPOSITION 3. *Independent Interim Rationalizability characterizes common certainty of rationality, the standard type space and independent beliefs. That is,*

(i) in any epistemic model, if $((e_i, e_{-i}), \theta) \in CC(Rat \cap W \cap Y)$*, then* $\mathrm{a}_i^*(e_i) \in IIR_i^*(\tau_i(e_i))$

(ii) if $a_i \in IIR_i^*(t_i)$*, then there exists an epistemic model and a state* $((e_i, e_{-i}), \theta)$ *such that* $((e_i, e_{-i}), \theta) \in CC(Rat \cap W \cap Y)$*,* $\tau_i(e_i) = t_i$*, and* $\mathrm{a}_i^*(e_i) = a_i$*.*

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 certainty of rationality and the type space—are needed to justify restricting attention to actions that are interim independent rationalizable on the type space. The additional assumption of common certainty 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.

The primitives of our model remain a finite set Θ of states of Nature, a finite set I of players, and a type space $\mathcal{T} = (T_i, \pi_i)_{i \in i}$. We now 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. For measurable X we denote by $\Delta(X)$ the set of (probability) measures on X.[18]

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

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

Each $\Delta(T_{-i} \times \Theta)$ gets the corresponding sigma algebra; we then assume that each $\pi_i : T_i \to \Delta(T_{-i} \times \Theta)$ 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.[19]

There is a belief-preserving morphism from one measurable type space into another measurable type space if the former space can be mapped into the latter while preserving the belief structure. Formally, there is a belief-preserving morphism from (T_i, π_i)

[18]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.

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

into $(\tilde{T}_i, \tilde{\pi}_i)$ if for each i there exists measurable $\varphi_i : T_i \to \tilde{T}_i$ with

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

for all measurable $Z \subset \tilde{T}_{-i} \times \Theta$. We call $\varphi = (\varphi_1, \dots, \varphi_n)$ 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 $(\delta^1, \delta^2, \dots) \in (\Delta(X^k))_{k=0}^{\infty} \triangleq H$ 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 $\mathcal{T}^* = (T_i^*, \pi_i^*)_{i\in I}$ that comprises a subset of hierarchies, $T_i^* \subset H$, and a measurable belief function, $\pi_i^* : T_i^* \to \Delta(T_{-i}^* \times \Theta)$, 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 i for notational simplicity. The type space is 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}-order 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 Θ are defined pointwise by

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

For each $k = 2, 3, \dots$ and measurable $Z \subseteq X^{k-1}$, let

$$\widehat{\pi}_i^k(t_i)[Z] = \pi_i(t_i)[\{(t_{-i},\theta) : (\widehat{\pi}_{-i}^1(t_{-i}), \dots, \widehat{\pi}_{-i}^{k-1}(t_{-i}), \theta) \in Z\}];$$

the morphism guarantees that $\widehat{\pi}_i^k : T_i \to T^{*,k}$ is measurable for each k. Let $\widehat{\pi}_i^*(t_i) = (\widehat{\pi}_i^k(t_i))_{k=1}^{\infty}$. Then $\widehat{\pi}_i^* : T_i \to T_i^*$ is the morphism φ_i discussed above.

We use the topology-free approach because we do not want to restrict ourselves to a particular topology and it enables us to provide stronger results, as they apply to all measurable type spaces. However, we rely on the fact that in our context there is a belief-preserving isomorphism between the universal type space $(T_i^*, \pi_i^*)_{i\in I}$ discussed above and the more familiar constructions using topological methods due to Mertens and Zamir (1985) (see also Brandenburger and Dekel 1993 and Heifetz 1993). These authors construct a universal type space $T^{**} = (T_i^{**}, \pi_i^{**})_{i\in I}$ with a topology on T_i^{**} under which $\pi_i^{**} : T_i^{**} \to \Delta(T_{-i}^{**} \times \Theta)$ are continuous and under which T_i^{**} are compact. They show this type space is universal for all continuous type spaces in the sense that for any type space $T = (T_i, \pi_i)_{i\in i}$ for which π_i is continuous according to a topology on T_i, there is a continuous belief-preserving morphism into T^{**}. Mertens et al. (1994, Theorem 1.3) show that in fact T^{**} is universal for all measurable type spaces—there is a belief-preserving morphism of any measurable type space into T^{**}. Given uniqueness (up to belief-preserving isomorphisms) of the universal type space constructed by Heifetz and Samet (1998, Proposition 3.5), there is a belief-preserving isomorphism between T^* and

T^{**}; we use this equivalence extensively in the proofs below and for notational simplicity write T^* for both. Even though the proofs therefore use the continuity properties of T^{**}, because of the isomorphism our results do not involve a topology or continuity.

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 conjectures over opponents' strategies are restricted to those actions. We write $\int_{T_{-i}} f(\cdot)\nu(\cdot)[(\mathrm{d}t_{-i},\theta,a_{-i})]$ when integrating with respect to t_{-i} only, holding θ and a_{-i} fixed.

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

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

Next we define

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

REMARK 4. Because A_i is finite, and expected 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 write $\Psi_i^{\mathcal{T}}(t_i,F)$ for the set of beliefs on the finite set $A_{-i}\times\Theta$ that are consistent with type t_i's beliefs and certainty that other players are choosing actions consistent with F_{-i}. Thus

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

This is the set of distributions over $A_{-i}\times\Theta$ 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

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

4.2.2 *Iterative definitions* We now define rationalizability as the result of iterating the BR map. As in the finite case, let $R_0^{\mathcal{T}}=((A_i)_{t_i\in T_i})_{i\in I}$, $R_k^{\mathcal{T}}=BR^{\mathcal{T}}(R_{k-1}^{\mathcal{T}})$, and $R^{\mathcal{T}}=\cap_{k=1}^{\infty}R_k^{\mathcal{T}}$. The corresponding objects on the universal type space are $R_0^*=((A_i)_{t_i^*\in T_i^*})_{i\in I}$, $R_k^*=BR^{\mathcal{T}^*}(R_{k-1}^*)$, 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, $R_{i,k}^{\mathcal{T}}(t_i)=R_{i,k}^*(t_i^*)$, $\Psi_i^{\mathcal{T}}(t_i,R_{-i,k}^{\mathcal{T}})=\Psi_i^*(\varphi_i(t_i),R_{-i,k}^{\mathcal{T}^*})$, $R_{i,k}^{\mathcal{T}}:T_i\to 2^{A_i}\setminus\emptyset$ is a measurable function, and $\{t_i^*\in T_i^*:a_i\in R_{i,k}^*(t_i^*)\}$ 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 $k'\leq k-1$. So suppose that for all i, and $t_i\in T_i$, $R_{i,k-1}^{\mathcal{T}}(t_i)=R_{i,k-1}^*(\varphi(t_i))$ and $\Psi_{i,k-1}^{\mathcal{T}}(t_i)=\Psi_{i,k-1}^*(\varphi(t_i))$, that $R_{i,k-1}^{\mathcal{T}}:T_i\to 2^{A_i}\setminus\emptyset$ is a measurable function, and that $\{t_i^*:a_i\in R_{i,k-1}^*(t_i^*)\}\subset T_i^*$ is closed.

(Part I) The set $\{t_i^*:a_i\in R_{i,k}^*(t_i^*)\}$ is closed and therefore measurable. To see this, consider a sequence $(t_i^{*n})_{n=1}^{\infty}$ that converges to t_i^* and such that $a_i\in R_{i,k}^*(t_i^{*n})$ for all n. Then for each t_i^{*n} there exists $\psi_i^{k-1,n}\in\Psi_{i,k-1}^*(t_i^{*n})$ such that

$$a_i\in\arg\max_{a_i'}\sum_{\theta,a_{-i}}g_i(a_i',a_{-i},\theta)\psi_i^{k-1,n}[(a_{-i},\theta)].$$

Moreover,

$$\psi_i^{k-1,n}[(a_{-i},\theta)]=\int_{t_{-i}^*}\sigma_{-i}^{*n}(t_{-i}^*,\theta)[a_{-i}]\pi_i^*(t_i^{*n})[(\mathrm{d}t_{-i}^*,\theta)]$$

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 $\nu^{*n}=\nu(\sigma_{-i}^{*n},\pi_i^*(t_i^{*n}))$, and by the compactness of $\Delta(T_{-i}^*\times\Theta\times A_{-i})$ 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 $\nu^{*\infty}=\nu(\pi_i(t_i^*),\sigma_{-i}^{*\infty})$.

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

$$a_i\in\arg\max_{a_i'}\sum_{\theta,a_{-i}}g_i(a_i',a_{-i},\theta)\psi_i[a_{-i},\theta].$$

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

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

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

Equation (1) can be written as $\pi^*_i(t^*_i)[N]=0$, where

$$N\equiv\{(t^*_{-i},\theta) :\ \operatorname{supp}\sigma^{*\infty}_{-i}(t^*_{-i},\theta)\not\subset R^*_{-i,k-1}(t^*_{-i})\}.$$

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

(Part II) 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 $R^{\mathcal{T}}_{i,k} : T_i\to 2^{A_i}\setminus\emptyset$ is measurable.

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

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

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

$$\psi_i[(\theta,a_{-i})]=\int_{t_{-i}}\sigma_{-i}(t_{-i},\theta)[a_{-i}]\pi_i(t_i)[(\mathrm{d}t_{-i},\theta)]$$

is in $\Psi^{\mathcal{T}}_i(t_i,R^{\mathcal{T}}_{-i,k})$.

From the morphism we have

$$\begin{aligned}\psi^*_i[(\theta,a_{-i})]&=\int_{t^*_{-i}}\sigma^*_{-i}(t^*_{-i},\theta)[a_{-i}]\pi^*_i(\varphi(t_i))[(\mathrm{d}t^*_{-i},\theta)]\\&=\int_{t_{-i}}\sigma_{-i}(t_{-i},\theta)[a_{-i}]\pi_i(t_i)[(\mathrm{d}t_{-i},\theta)].\end{aligned}$$

Thus

$$\Psi^*_{i,k}(t^*_i)\subset\Psi^{\mathcal{T}}_{i,k}(t_i).$$

(Part IVb) To prove the converse, suppose $\psi_i \in \Psi_i^{\mathcal{T}}(t_i, R_{-i,k}^{\mathcal{T}})$ and let σ_{-i} be the associated conjecture so $\psi_i[(\theta, a_{-i})] = \int_{t_{-i}} \sigma_{-i}(t_{-i}, \theta)[a_{-i}]\pi_i(t_i)[(\mathrm{d}t_{-i}, \theta)]$. We define $\sigma_{-i}^* : T_{-i}^* \times \Theta \to \Delta(A_{-i})$ as follows.

First, for every $B_{-i} \subset A_{-i}$ let $\tau_{-i}^{\mathcal{T}}(B_{-i}) = \{t_{-i} \in T_{-i} : B_{-i} = R_{-i,k-1}^{\mathcal{T}}(t_{-i})\}$ and $\tau_{-i}^*(B_{-i}) = \{t_{-i} \in T_{-i}^* : B_{-i} = R_{-i,k-1}^*(t_{-i})\}$. By the induction hypothesis both $\tau_{-i}^{\mathcal{T}}(B_{-i}) \subset T_{-i}$ and $\tau_{-i}^*(B_{-i}) \subset T_{-i}^*$ are measurable and $\tau_{-i}^{\mathcal{T}}(B_{-i}) = \varphi_{-i}^{-1}(\tau_{-i}^*(B_{-i}))$, and hence $\pi_i^*(\varphi(t_i))[\tau_{-i}^*(B_{-i})] = \pi_i(t_i)[\tau_{-i}^{\mathcal{T}}(B_{-i})]$.

We construct $\sigma_{-i}^*(t_{-i}^*, \theta)[\cdot] \in \Delta(A_{-i})$ as follows. Map $\sigma_{-i}(t_{-i}, \cdot)$ into $\sigma_{-i}^*(t_{-i}^*, \cdot)$ by taking all t_{-i} for whom B_{-i} is $k-1$ rationalizable, denoted $\tau_{-i}^{\mathcal{T}}(B_{-i})$, taking the conditional average of $\sigma_{-i}(t_{-i}, \cdot)$ over those t_{-i}, and assigning that average conjecture to those t_{-i}^* who have that same $k-1$ rationalizable set, i.e., to $\tau_{-i}^*(B_{-i})$. Moreover, each $\tau_{-i}^*(B_{-i})$ is a superset of $\varphi(\tau_{-i}^{\mathcal{T}}(B_{-i}))$, and these supersets 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 is not well defined because the conditioning event, $\tau_{-i}^{\mathcal{T}}(B_{-i})$, 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 $t_{-i}^* \in \tau_{-i}^*(B_{-i})$ whenever $\pi_i(t_i)[\tau_{-i}^{\mathcal{T}}(B_{-i})] = 0$. To do this, for each B_{-i} fix some $\bar{a}_{-i}(B_{-i}) \in B_{-i}$.

We now formalize this verbal description. Let

$$\sigma_{-i}^*(t_{-i}^*, \theta)[a_{-i}] = \begin{cases} \dfrac{\int_{\tau_{-i}^{\mathcal{T}}(B_{-i})} \sigma_{-i}(t_{-i}, \theta)[a_{-i}]\pi(t_i)[(\mathrm{d}t_{-i}, \theta)]}{\pi(t_i)[\tau_{-i}^{\mathcal{T}}(B_{-i})]} & \text{if } t_{-i}^* \in \tau_{-i}^*(B_{-i}) \text{ and } \pi(t_i)[\tau_{-i}^{\mathcal{T}}(B_{-i})] > 0 \\ 1 & \text{if } t_{-i}^* \in \tau_{-i}^*(B_{-i}),\ \pi(t_i)[\tau_{-i}^{\mathcal{T}}(B_{-i})] = 0 \text{ and } a_{-i} = \bar{a}_{-i}(B_{-i}) \\ 0 & \text{if } t_{-i}^* \in \tau_{-i}^*(B_{-i}),\ \pi(t_i)[\tau_{-i}^{\mathcal{T}}(B_{-i})] = 0 \text{ and } a_{-i} \neq \bar{a}_{-i}(B_{-i}). \end{cases}$$

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

Now

$$\begin{aligned} \psi_i^*[\theta, a_{-i}] &= \int_{t_{-i}^*} \sigma_{-i}^*(t_{-i}^*, \theta)[a_{-i}]\pi_i^*(\varphi_i(t_i))[(\mathrm{d}t_{-i}^*, \theta)] \\ &= \sum_{B_{-i} \subset A_{-i}} \left(\frac{\int_{\tau_{-i}^{\mathcal{T}}(B)} \sigma_{-i}(t_{-i}, \theta)[a_{-i}]\pi_i(t_i)[(\mathrm{d}t_{-i}, \theta)]}{\pi_i(t_i)[\tau_{-i}^{\mathcal{T}}(B_{-i})]} \right) \pi_i^*(\varphi(t_i))[\tau_{-i}^*(B_{-i}) \times \{\theta\}] \\ &= \sum_{B_{-i} \subset A_{-i}} \left(\frac{\int_{\tau_{-i}^{\mathcal{T}}(B)} \sigma_{-i}(t_{-i}, \theta)[a_{-i}]\pi_i(t_i)[(\mathrm{d}t_{-i}, \theta)]}{\pi_i(t_i)[\tau_{-i}^{\mathcal{T}}(B_{-i})]} \right) \pi_i(t_i)[\tau_{-i}^{\mathcal{T}}(B_{-i}) \times \{\theta\}] \\ &= \int_{t_{-i}} \sigma_{-i}(t_{-i}, \theta)[a_{-i}]\pi_i(t_i)[(\mathrm{d}t_{-i}, \theta)] = \psi_i[\theta, a_{-i}], \end{aligned}$$

where the first equality is by definition, the second by substitution and changing the integration of a finite valued "step function" to a sum, the third using $\pi_i^*(\varphi(t_i))[\tau_{-i}^*(B_{-i})] =$

$\pi(t_i)[\tau^{\mathscr{T}}_{-i}(B_{-i})]$, and the last changing the summation of the "step function" back to an integral. Thus for any $\psi \in \Psi^{\mathscr{T}}_i(t_i, R^{\mathscr{T}}_{-i,k})$ we have found $\psi^* \subset \Psi^*_i(t^*_i, R^*_{-i,k})$. □

Corollary 2. *$R^{\mathscr{T}}_i(t_i) = R^{\mathscr{T}'}_i(t'_i)$ if $\widehat{\pi}^*_i(t_i) = \widehat{\pi}^*_i(t'_i)$.*

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 as in the finite case. The properties mentioned there also hold, although the argument is slightly different.

Lemma 2. *(i) If $S^{\mathscr{T}}_c$ for all c in some index set C are best-reply sets then $\cup_c S^{\mathscr{T}}_c$ is a best-reply set.*

(ii) The union of all best-reply sets is the largest fixed point of $BR^{\mathscr{T}}$.

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

Definition 4. $R^{\mathscr{T}}_{\mathscr{F}} = ((R^{\mathscr{T}}_{i,\mathscr{F}}(t_i))_{t_i \in T_i})_{i \in I} \subset ((2^{A_i})^{T_i})_{i \in I}$ is the largest fixed point of $BR^{\mathscr{T}}$.

In general, the largest fixed point need not 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. *$R^{\mathscr{T}}_{\mathscr{F}}$ equals $R^{\mathscr{T}}$.*

Proof. It is sufficient to prove that $R^{\mathscr{T}}$ is a best-reply set. That nothing larger can be a best-reply set is immediate. For every $a_i \in R^{\mathscr{T}}_i(t_i)$ we have that for every k there is a measurable $\sigma^k_{-i} : T_{-i} \times \Theta \to \Delta(A_{-i})$ such that $\sigma^k_{-i}(t_{-i}, \theta)[a_{-i}] > 0 \Rightarrow a_j \in R^{\mathscr{T}}_{k,t_j}(t_j)$ and

$$a_i \in \arg\max_{a'_i} \sum_{\theta, a_{-i}} \int_{t_{-i}} g_i(a'_i, a_{-i}, \theta) \sigma^k_{-i}(t_{-i}, \theta)[a_{-i}] \pi(t_i)[(\mathrm{d}t_{-i}, \theta)].$$

We need to prove there exists $\sigma_{-i} : T_{-i} \times \Theta \to \Delta(A_{-i})$ such that $\sigma_{-i}(t_{-i}, \theta)[a_{-i}] > 0 \Rightarrow a_j \in \cap_k R^{\mathscr{T}}_{k,t_j}(t_j)$ and

$$a_i \in \arg\max_{a'_i} \sum_{\theta, a_{-i}} \int_{t_{-i}} g_i(a'_i, a_{-i}, \theta) \sigma_{-i}(t_{-i}, \theta)[a_{-i}] \pi(t_i)[(\mathrm{d}t_{-i}, \theta)].$$

Define $\sigma_{-i}(t_{-i}, \theta) = \lim_k \sigma^k_{-i}(t_{-i}, \theta)$. We need only check then that σ_{-i} is measurable. But

$$\{t_i : \sigma_{-i}(t_{-i}, \theta) = a_{-i}\} = \bigcup_{K=1}^{\infty} \bigcap_{k \geq K} \{t_i : \sigma^k_{-i}(t_{-i}, \theta) = a_{-i}\};$$

since the latter is measurable so is the former. Given two type spaces, $\mathscr{T}$ and $\mathscr{T}'$, on the set of states of Nature Θ, with t_i a type of i in $\mathscr{T}$ and t'_i a type of i in $\mathscr{T}'$, we have $\widehat{\pi}^*_i(t'_i) = \widehat{\pi}^*_i(t_i) \Rightarrow R^{\mathscr{T}'}_{i,\mathscr{F}}(t'_i) = R^{\mathscr{T}}_{i,\mathscr{F}}(t_i)$. □

5. Concluding remarks

Remark 5. Weinstein and Yildiz (forthcoming) show that the set of types in the universal type space with a unique ICR action is open and dense in the product topology. Dekel et al. (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.

Remark 6. Despite the difference between ICR and IIR highlighted above, these concepts are equivalent in some games. While it is beyond the scope of this paper to provide details, it can be shown that these include games where there are strategic complementarities, two-person private-value games where each player knows her value, and of course whenever ICR delivers a unique outcome.

Remark 7. The ICR and IIR solution concepts are interim concepts in the sense that a type's conjecture about the play of opposing players is specified as a function of his type, so that the conjectures of different types of the same player can be different. Under an "ex ante" approach each player i has a prior determined before learning her own type, leading to differences with interim concepts that are analogous to the difference between the ex ante and interim versions of dominance that was pointed out by Fudenberg and Tirole (1991). Indeed, if one considers rationalizability notions where a player's conjecture σ_{-i} about others' play is not allowed to depend on his type, then such ex ante notions analogously differ from the interim notions we define. However as a referee pointed out, an alternative approach to an ex ante definition of correlated rationalizability would specify, instead of a belief for each type, $\nu_{t_i} \in \Delta(T_{-i} \times \Theta \times A)$, a prior for each player, $\mu \in \Delta(T \times \Theta \times A)$. If the prior allows for correlation then ν_{t_i} are the conditionals given t_i of some prior μ, i.e., $\mu(\cdot \mid t_i) = \nu_{t_i}(\cdot)$, and hence with correlation the ex ante notion does not impose the restriction that beliefs of different types of the same player coincide, as would be implied in the cases discussed above or with independence assumptions.[20] However, such an ex ante definition would still differ from ICR because the interim perspective in the latter does serve as a refinement (similar to extensive-form rationalizability (Pearce 1984)) since zero probability types are required to optimize under ICR but not under the ex ante solution concepts defined above. For example, if player 1's ex ante probability of a given type t_1' is 0, while player 2 assigns it positive probability, the ex ante concept would let type t_1' play an interim dominated strategy, which could allow player 2 to play an action that would be ruled out by ICR.

References

Aumann, Robert J. (1987), "Correlated equilibrium as an expression of Bayesian rationality." *Econometrica*, 55, 1–18. [24, 27, 30]

Aumann, Robert J. and Adam Brandenburger (1995), "Epistemic conditions for Nash equilibrium." *Econometrica*, 63, 1161–1180. [27]

[20] As noted, if independence is imposed, then in the ex ante notion the conditionals are constant across t_i's while the interim notion allows for different forecasts for each t_i, so the two differ.

Battigalli, Pierpaolo and Marciano Siniscalchi (2003), "Rationalization and incomplete information." *B. E. Journal of Theoretical Economics*, 3, issue 1 (Advances), article 3. [15, 19]

Bergemann, Dirk and Stephen Morris (2005), "Robust mechanism design." *Econometrica*, 73, 1771–1813. [15]

Bernheim, B. Douglas (1984), "Rationalizable strategic behavior." *Econometrica*, 52, 1007–1028. [24]

Brandenburger, Adam and Eddie Dekel (1987), "Rationalizability and correlated equilibria." *Econometrica*, 55, 1391–1402. [16, 24, 27, 31]

Brandenburger, Adam and Eddie Dekel (1993), "Hierarchies of beliefs and common knowledge." *Journal of Economic Theory*, 59, 189–198. [32]

Brandenburger, Adam and Amanda Friedenberg (2006), "Intrinsic correlation in games." Unpublished paper, Washington University, St. Louis. [25]

Dekel, Eddie, Drew Fudenberg, and Stephen Morris (2006), "Topologies on types." *Theoretical Economics*, 1, 275–309. [16, 24, 38]

Ely, Jeffrey C. and Marcin Pęski (2006), "Hierarchies of belief and interim rationalizability." *Theoretical Economics*, 1, 19–65. [19, 24]

Forges, Françoise (1993), "Five legitimate definitions of correlated equilibrium in games with incomplete information." *Theory and Decision*, 35, 277–310. [19]

Fudenberg, Drew and Jean Tirole (1991), *Game Theory*. MIT Press, Cambridge, Massachusetts. [38]

Harsanyi, John C. (1967–68), "Games with incomplete information played by 'Bayesian' players, Parts I, II, III." *Management Science*, 14, 159–182, 320–334, and 486–502. [15]

Heifetz, Aviad (1993), "The Bayesian formulation of incomplete information—the non-compact case." *The International Journal of Game Theory*, 21, 329–338. [32]

Heifetz, Aviad and Dov Samet (1998), "Topology-free typology of beliefs." *Journal of Economic Theory*, 82, 324–341. [17, 31, 32]

Lipman, Barton L. (1994), "A note on the implications of common knowledge of rationality." *Games and Economic Behavior*, 6, 114–129. [37]

Mertens, Jean-François, Sylvain Sorin, and Shmuel Zamir (1994), "Repeated games, Part A: Background material." Discussion Paper 9420, Center for Operations Research and Econometrics, Catholic University of Louvain. [32]

Mertens, Jean-François and Shmuel Zamir (1985), "Formulation of Bayesian analysis for games with incomplete information." *International Journal of Game Theory*, 14, 1–29. [15, 18, 19, 32]

Pearce, David G. (1984), "Rationalizable strategic behavior and the problem of perfection." *Econometrica*, 52, 1029–1050. [24, 26, 38]

Tan, Tommy Chin-Chiu and Sérgio Ribeiro da Costa Werlang (1988), "The Bayesian foundations of solution concepts of games." *Journal of Economic Theory*, 45, 370–391. [27]

Weinstein, Jonathan and Muhamet Yildiz (forthcoming), "A structure theorem for rationalizability with application to robust predictions of refinements." *Econometrica*. [38]

Submitted 2005-9-6. Final version accepted 2006-12-4. Available online 2006-12-5.