

# An Epistemic Characterization of Extensive Form Rationalizability

Pierpaolo Battigalli  
Princeton University

Marciano Siniscalchi  
Stanford University

## JOB MARKET PAPER

August 1997\*

### Abstract

We use an extensive form, universal type space to provide the following epistemic characterization of extensive form rationalizability. Say that player  $i$  *strongly believes* event  $E$  if  $i$  is certain of  $E$  conditional on each of her information sets consistent with  $E$ . Our main contribution is to show that a strategy profile  $s$  is extensive form rationalizable if and only if there is a state in which  $s$  is played and (0) everybody is rational, (1) everybody strongly believes (0), (2) everybody strongly believes (0) & (1), (3) everybody strongly believes (0) & (1) & (2),  $\dots$ . This result also allows us to provide sufficient epistemic conditions for the backward induction outcome and to relate extensive form rationalizability and conditional common certainty of rationality.

## 1. Introduction

Extensive-form rationalizability (Pearce [17], Battigalli [5], [7]) attempts to capture the implications of rationality and common certainty of rationality in extensive games, and incorporates a powerful, yet quite natural notion of forward induction (see especially [5]).

---

\*Paolo Battigalli worked on this paper while he was visiting the Division of the Humanities and Social Sciences at Caltech. We thank Elchanan Ben-Porath, Eddie Dekel, Bob Wilson and seminar participants at Caltech, Stanford and Tel Aviv University for helpful comments and suggestions. All remaining errors are our own. Email: [battigal@phoenix.princeton.edu](mailto:battigal@phoenix.princeton.edu), [marciano@leland.stanford.edu](mailto:marciano@leland.stanford.edu).

Its intuitive justification is thus similar to that leading to normal-form rationalizability (Pearce [17], Bernheim [10]). However, while epistemic characterizations of the latter solution concept have been provided (e.g. Tan and Werlang [25]), no such result exists for extensive-form rationalizability. The main purpose of this paper is to fill this void.

To obtain such a characterization, one essentially needs two key ingredients: an epistemic model which can capture the subtleties of extensive-form reasoning, and an axiom system which correctly embodies the intuitive underpinnings of extensive-form rationalizability.

In this paper we use an extensive form epistemic model based on Alfred Renyi's and Roger Myerson's conditional probability systems (see [20], [15]), developed by Battigalli [6]. This is presented and briefly discussed in Section 2.2. The key idea here is that players may update their beliefs as the game progresses. Furthermore, players need to form conjectures on their opponents' future play, as well as on their opponents' rationality and epistemic status. A player's *type* is thus essentially an *infinite hierarchy of conditional beliefs*. Battigalli's model allows to capture these elements, while imposing a natural (and convenient) Bayesian consistency restriction on beliefs. A key feature of this epistemic model is that it is *universal*: every "conceivable" profile of hierarchies of conditional beliefs is included in the model. This is especially important in order to correctly model the players' *forward induction* reasoning. Essentially, at each point in the game, a player who applies forward induction looks for the opponent's epistemic types and strategies that rationalize the observed behavior. This search is necessarily limited to types within the postulated epistemic model; thus, unless the latter is rich enough, i.e includes every conceivable epistemic type, substantive restrictions are placed on the players' inferences. In this case, the forward induction principle may potentially lose its bite, or lead to arbitrary conclusions which crucially depend on the specific restrictions embodied in the postulated type space. Of course, one may take the point of view that, in a given situation, certain restrictions are actually desirable; however, our goal in this paper is to provide a "neutral" analysis of the forward induction logic as embodied in extensive-form rationalizability. Therefore, we avoid extraneous restrictions on beliefs.<sup>1</sup>

The axiom system (see Section 3.2) incorporates two key ideas. First, the notion of "strong belief" plays a crucial role. Essentially, we say that a player *strongly believes* that an event  $E$  is true if she assigns (conditional) probability one to  $E$  at every information set

---

<sup>1</sup>Moreover, we argue that adopting a universal epistemic model allows the modeller to be quite explicit (and transparent) about exogenous restrictions on beliefs. Indeed, once such restrictions are made explicit, they may be combined with forward induction reasoning coherently and (in our opinion) rather naturally.

which is not inconsistent with  $E$  having occurred.

The second ingredient is the *best rationalization principle*: that is, the idea that, at each point in the game, players bestow the highest possible degree of strategic sophistication upon their opponents. This is the forward-induction assumption discussed in Battigalli [5] and [7].

The main representation result, Proposition 4.4, shows that, *in any given extensive-form game, a strategy profile  $s$  survives  $n+1$  rounds of extensive form rationalizability if and only if there is a state where  $s$  is played and all the following events are true*

- (0) every player is rational,
- (1) every player strongly believes (0),
- (2) every player strongly believes (0) & (1),
- .....
- ( $n$ ) every player strongly believes (0) & (1) & ... & ( $n-1$ ).

Since extensive form rationalizability is generically outcome-equivalent to backward induction [7], we obtain as a corollary the following result: Consider a generic  $N$ -person game of perfect information where each player has at most  $(k+1)$  reduced normal form strategies: in every state where the above events (0), (1), ..., ( $n$ ) are true the players follow the backward induction path. Another consequence of the main result is that, in every multistage game with observable actions, there can be common certainty of rationality conditional on partial history  $h$  if there is a profile of extensive form rationalizable strategies reaching  $h$ . Finally, Tan and Werlang's characterization of correlated rationalizability (iterated strict dominance) also follows from our main result as a special case.

This paper builds on previous work independently conducted by Battigalli [6] and Siniscalchi [22]. Our extensive form epistemic model can be regarded as a generalization of the model used by Ben Porath [9] to characterize common certainty of rationality at the beginning of a perfect information game. Stalnaker [23] and [24] consider a related normal form model, which can also be used to analyze extensive form reasoning. Unlike our epistemic model, those of Ben Porath and Stalnaker are not universal (for more on this comparison see [6]). [24] puts forward a notion of "robust belief" which corresponds to our "strong belief" and briefly discusses the relation between robust belief in rationality and forward induction.<sup>2</sup>

---

<sup>2</sup>The relationship between Stalnaker [24] and our paper is as follows. An epistemic model *a la* Stalnaker induces a finite, belief-closed subspace of our universal type space and his notion of perfect rationality is equivalent to our notion of weakly sequential rationality in games with generic payoffs at terminal nodes. Stalnaker shows that a strategy profile  $s$  is realized at some state  $w$  of any "sufficiently rich" epistemic model

Aumann [2], [3], [4] and Samet [21] use different epistemic models and provide a different set of sufficient conditions for the backward induction outcome. Their results involve the notion of common knowledge rather than common certainty<sup>3</sup> and do not contain explicit assumptions concerning how players revise their beliefs when they face unexpected evidence (although Samet comes somewhat closer to this with his notion of "hypothetical knowledge"). Finally, in the context of a partitional model, Asheim and Dufwemberg [1] formalize the notion of "common certainty of admissibility" and thereby characterize an iterated deletion procedure which captures certain aspects of forward induction.

The paper is structured as follows. Section 2 presents two examples which illustrate the logic of extensive-form rationalizability and its connections with backward and forward induction, the role of epistemic types, the consequences of adopting *ad hoc*, non-universal models, and the notion of strong belief. In Section 3 we lay down the game-theoretic, epistemic model proper. In particular, Subsection 3.2 contains the construction of our extensive-form, universal type space. In Section 4, we define extensive form rationalizability and strong beliefs and we state the characterization result. As a by-product, we can provide sufficient epistemic conditions for the backward induction outcome and relate extensive form rationalizability to conditional common certainty of rationality. All proofs are in the Appendix.

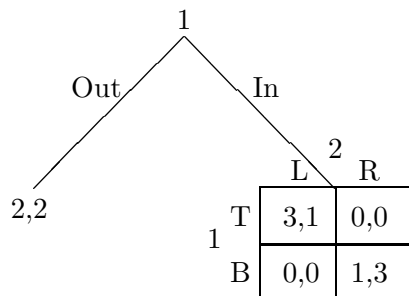
---

$M$  where there is common certainty (at the outset) of perfect rationality and of the fact that every player strongly believes in rationality if and only if  $s$  survives two rounds of iterative weak dominance followed by arbitrarily many rounds of iterative strong dominance. He mentions that similar results involving more rounds of iterative weak dominance could be proved. Note that  $n$  rounds of iterative weak dominance are generically equivalent to  $n$  rounds of extensive form rationalizability (see Battigalli [7]) therefore such results are related to our main proposition. The key difference between Stalnaker's approach and ours is that he does not work with a universal model. This implies that a result involving  $n$  rounds of weak dominance holds for an *ad hoc* class of "sufficiently rich" models, so that there are "enough types" to be able to rationalize the observed behavior of the opponents at information sets consistent with  $k = 1, \dots, n - 1$  rounds of iterative weak dominance.

<sup>3</sup>In axiomatic characterizations of the knowledge operator (e.g. S5), one assumes that it is not the case that an agent *knows* a false proposition. On the other hand, axiom systems for the certainty operator (e.g. KD45) encompass situations in which an agent may be *certain* that a given proposition is true, although in reality it is (or subsequently turns out to be) false (details may be found in [14]). For precisely this reason, we feel that the notion of certainty is more germane to extensive-form analysis than the stronger notion of knowledge.

## 2. Two Examples

The following game is a variant of the well-known “Battle of the Sexes” with an outside option:



Notice that the strategy profiles  $((\text{Out}, \text{B}), \text{R})$  and  $((\text{In}, \text{T}), \text{L})$  are proper (hence sequential) equilibria of this game. The former is supported by Player 2’s off-equilibrium-path belief that Player 1 would follow In with B. However, the usual forward induction argument goes, this belief is not “reasonable” because the strategy  $(\text{In}, \text{B})$  is strictly dominated by  $(\text{Out}, *)$  for Player 1; hence, if he gets a chance to move, Player 2 should infer that Player 1 plans to play T, not B, so that his best reply is to play L. But then Player 1, anticipating all this, will not want to exit the game at her first information node.

The argument relies on a restriction on Player 2’s beliefs (off the equilibrium path) concerning Player 1’s rationality, and on a restriction on Player 1’s beliefs about Player 2’s beliefs. In order to formalize this kind of reasoning, it seems natural to introduce the notion of a player’s *epistemic type*, in the spirit of John Harsanyi’s analysis of incomplete information games. In keeping with forthcoming notation, we define a set of possible “states of the world” comprising an epistemic type *and* a strategy for each player (see Ben-Porath [9]). In the case of the game we are considering, an epistemic type for Player 2 is (modulo a technicality which we shall discuss momentarily) a conditional probability distribution on the opponent’s strategy-type pairs<sup>4</sup>, whereas an epistemic type for Player 1 should comprise *two* probability distributions on Player 2’s strategy-type pairs – one representing Player

---

<sup>4</sup>In multistage games with observed actions, we can assume that players have beliefs at *every* information set – even at those which they do not control. However, in most of the development to follow (including the main characterization result) we only assume that players have conjectures at each information set they own. Thus, in particular, we do not consider Player 2’s conjecture at Player 1’s initial node in this game.

1's beliefs at her first information node, and another representing her conjecture following her choice of In. One obvious restriction is that Player 2's beliefs should assign positive probability only to strategies prescribing the move In at Player 1's first information set. This should hold also at states where Player 1 chooses Out; in fact, a state also comprises the beliefs that players *would* hold conditional on *counterfactual* information sets – that is, information sets that the players' own planned moves in a given state actually *prevent* from being reached.

Consider for example the following strategy-type pairs (the first table refers to Player 1, while the second is for Player 2):

	Pair	$p_{11}$	$p_{21}$
1	((In,B), $t_{11}$ )	0,1	0,1
2	((In, T), $t_{21}$ )	0,1	0,1
3	((Out,B), $t_{31}$ )	0,1	0,1

	Pair	$p_2$
1	(L, $t_{12}$ )	0,1,0
2	(R, $t_{22}$ )	1,0,0

We can describe states of the world compactly using the indices of the relevant strategy-type pairs for each player, as listed in the tables above. Consider for example state (1,1). Player 1 expects Player 2 to play R; she is also certain that Player 2 will be certain (in the subgame) that her type is indeed  $t_{11}$ , and that she is playing (In,B). Player 1 is clearly not rational and has incorrect beliefs about player 2. On the other hand Player 2 is rational, but he has incorrect beliefs about player 1.

Observe that all types of Player 1 share the same beliefs over the set of strategy-type pairs of Player 2; yet, we regard these types as different. The reason has to do with the technicality alluded to above: strictly speaking, a player's epistemic type comprises beliefs over *her own strategies* as well; however, our axioms will require that these beliefs be correct. Although this is not explicitly indicated in the preceding tables, we assume that this is indeed the case for all types listed above; thus, for instance, type  $t_{11}$  of Player 1 assigns probability 1 to the strategy (In, B).<sup>5</sup>

---

<sup>5</sup>There is another subtlety which is perhaps worth noting. Type  $t_{31}$  of Player 1 is required to hold beliefs at her second information set, although the action Player 1 plans to take in states  $(3, j)$ , for  $j = 1, 2, 3$ , clearly preclude it from being reached. This is a consequence of the formal definition of a type. Observe in particular that type  $t_{31}$  does not revise her beliefs *about Player 2*. While this may seem plausible (after all, type  $t_{31}$ 's conjecture about Player 2's behavior was not falsified) our axioms do not impose any such restriction – although they do require that type  $t_{31}$  assign probability zero to Player 1's initial choice of “Out” if Player 1's second information set is reached. At any rate, such beliefs play no role in our analysis.

In a similar vein, we also implicitly assume that players are certain of their own type.<sup>6</sup>

The sequential equilibrium ((Out, B), R) corresponds to state (3,2). Player 2, upon being reached, must conclude that Player 1 played In; in particular, if Player 2 (counterfactually) had to move, he would believe that Player 1 follows In with B, so he would respond with R. Also, if Player 1 believes that Player 2 would play R, she does well to choose Out.

The key observation now is that no state is consistent with the forward induction story. The problem is that the epistemic model considered so far is *not rich enough*; for instance, we have ruled out “conceivable” states of the world in which Player 1 expects Player 2 to play L. Hence, we have “forced” Player 2 to believe that only “irrational types” of Player 1 choose In. This being the case, it is not surprising that we are unable to formalize the forward-induction argument proposed at the beginning of this section.

We can “solve” the problem by introducing additional strategy-type pairs, as follows:

	Pair	$p_{11}$	$p_{21}$
1	((In,B), $t_{11}$ )	0,1,0	0,1,0
2	((In,T), $t_{21}$ )	0,1,0	0,1,0
3	((Out,B), $t_{31}$ )	0,1,0	0,1,0
4	((In, T), $t_{41}$ )	0,0,1	0,0,1

	Pair	$p_2$
1	(L, $t_{12}$ )	0,1,0,0
2	(R, $t_{22}$ )	1,0,0,0
3	(L, $t_{32}$ )	0,0,0,1

The addition of types  $t_{41}$  and  $t_{32}$  makes it possible to formalize the forward-induction argument simply by tracing the restrictions imposed by the axioms. First, the only states in which both players are rational are of the form  $(i, j)$  for  $i = 3, 4$  and  $j = 1, 2, 3$ . Next, in states of the form  $(i, 1)$  and  $(i, 2)$  Player 2 does not believe, upon being reached, that Player 1 is rational – although there is one rational strategy-type pair for Player 1 which supports Player 1’s initial choice of “In”. Hence, only states (3,3) and (4,3) are consistent with axioms (0) “every player is rational” and (1) “every player strongly believes (0)”. In terms of the forward induction argument, axiom (1) forces Player 2 to seek a rational explanation for Player 1’s observed behavior. Finally, axiom (2), “every player strongly believes (0) & (1)”, eliminates (3,3): that is, Player 1 must expect Player 2 to seek a rational explanation for the observed behavior – which is again precisely what the forward induction argument implies. Thus, only state (4,3) is consistent with axioms (0), (1) and (2). It is easy to see that it is also consistent with the subsequent axioms.

---

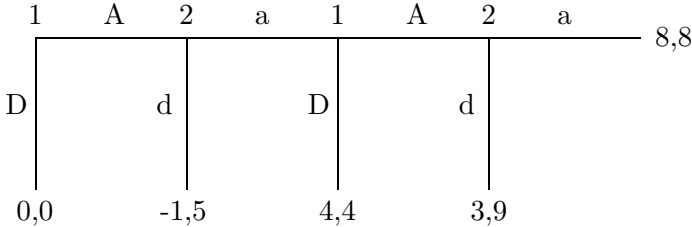
<sup>6</sup>Making this assumption explicit within our framework is possible but irrelevant for our purposes.

This is perhaps a good place to comment on the sense in which the original set of types is “embedded” in the second, enlarged space. The key idea is that there is (i) an obvious map carrying types of one player in the “old” space to the corresponding types of the same player in the “new” one, and (ii) an equally natural map carrying distributions over the set of “old” types of the opponent to the corresponding face of the simplex of distributions over the “new” opponent types. The crucial feature of these maps is that they *commute*: for instance, start with type  $t_{11}$  in the original type space, map it to the new  $t_{11}$ , and notice that this type assigns the probability distribution  $(0,1,0)$  over Player 2’s new types to Player 1’s second information set; equivalently, observe that type  $t_{11}$  in the old space assigns the distribution  $(0,1)$  over Player 2’s old types to the same information set, which indeed maps to the distribution  $(0,1,0)$  over Player 2’s new types.

In an entirely similar way, one can define an embedding of *any* prespecified extensive form type space in the *universal* type space which we shall construct shortly (see Battigalli [6] for details). The key feature of the universal space is that *every* (Bayes-consistent) array of conditional beliefs about the strategies and epistemic types of player  $j$  corresponds to a unique epistemic type of player  $i$ .

Thus, one can view any pre-specified type space as a particular subset of the universal type space. Types in this subset will exhibit an obvious property, called *belief-closedness*: at every information set, every player’s type in the subset (i) assigns probability one to types which lie in the subset; (ii) indeed, it assigns probability one to types which satisfy (i); (iii) it assigns probability one to types which satisfy (ii); etc. That is, at each information set there is common certainty that players’ types belong to the subset. Clearly, this property must necessarily hold for types in the originally specified space, so by construction it is inherited by their images in the universal space; again, see Battigalli [6] for details.

We conclude this section with an illustration of our result on backward induction in generic perfect information games. We choose an example (a variant of the centipede game) which also allows us to clarify the crucial notion of *strong belief*:





Consider the following epistemic model – equivalently, the following belief-closed subset of the universal model we will construct shortly:

	Pair	$p_{11}$	$p_{21}$		Pair	$p_{12}$	$p_{22}$
1	$((A,D), t_{11})$	1,0,0,0	1,0,0,0	1	$((a,a), t_{12})$	0,1,0,0	0,1,0,0
2	$((A,A), t_{21})$	$\frac{1}{8}, \frac{7}{8}, 0, 0$	1,0,0,0	2	$((d,*), t_{22})$	1,0,0,0	0,1,0,0
3	$((A,D)), t_{31})$	0,0,1,0	0,0,1,0	3	$((a,d), t_{32})$	0,1,0,0	0,1,0,0
4	$((D,*), t_{41})$	0,0,0,1	0,0,1,0	4	$((d,*), t_{42})$	0,0,1,0	0,1,0,0

We continue to assume that all player types hold correct beliefs concerning their own strategy in a given state of the world.

Notice that now both players' conjectures comprise a *pair* of probability distributions over the opponent's strategy-type pairs; however, we do not specify moves at information nodes which are precluded by the previous actions of their owner.

Let us begin by noting that the event, “All players are rational”, corresponding to axiom (0), characterizes the set  $\{(i, j) : i, j \in \{2, 3, 4\}\}$ . In particular, in any state  $(1, j)$ , Player 1's second choice is not rational if she expects Player 2 to choose  $(a,a)$  –  $(A,A)$  does strictly better; Player 2's strategy in any state  $(i, 1)$  is also irrational.

Next, it is easy to see by inspecting the supports of the probabilities  $p_{i1}$  and  $p_{j2}$  that the event (1) = “All players strongly believe (0)” corresponds to  $\{(i, j) : i \in \{3, 4\}, j \in \{3, 4\}\}$ . Notice that Player 2's type  $t_{12}$  believes that Player 1 is rational, but since we assume that Player 2 is certain of his strategy-type, we must conclude that he is also certain of his own irrationality. Also, notice that we must rule out Player 1's type  $t_{21}$ : although ex-ante she attaches positive probability to Player 2's rationality, this probability is not 1; also, if her second node is reached, by Bayesian consistency she is forced to believe with probability 1 that Player 2 is in fact irrational.

All told, (0) & (1) corresponds to  $\{(i, j) : i, j \in \{3, 4\}\}$ . Hence, only types  $t_{31}$  and  $t_{41}$  of Player 1, and type  $t_{42}$  of Player 2 strongly believe (0) & (1). The argument for  $t_{42}$  is subtle but important to grasp: at his second information node, he attaches probability 1 to the negation of (0) & (1); however, he cannot avoid doing this, because if (0) & (1) were true, his second node would never be reached! This is essentially what we mean by *strong belief*: it is a restriction which applies whenever it is not contradicted by the evidence.

Now (0) & (1) & (2) corresponds to  $\{(i, 4) : i \in \{3, 4\}\}$ ; thus, applying axiom (3), we are left with state (4,4) only. Again, note that Player 1's type  $t_{41}$  qualifies because her second

node cannot be reached if (0) & (1) & (2) is true.

Finally, note that also (0) & (1) & (2) & (3) corresponds to state (4,4); again,  $t_{41}$  will satisfy Axiom (4), *and so will*  $t_{42}$ , again by the notion of strong belief. Successive iterations will not impose any additional restrictions.

Now notice that state (4,4) yields the (unique) backward induction path, in accordance with our results.<sup>7</sup> Observe also that the subset  $\{(i, j) : i, j \in \{1, 2, 3\}\}$  is belief-closed, but does not include the “backward/forward induction” state of the world. This is a further indication of the relevance of a priori restrictions on the epistemic types of the players.

### 3. Game-Theoretic setup and Epistemic Model

#### 3.1. Extensive Form Games

For simplicity we consider finite extensive games with complete (but possibly imperfect) information, perfect recall and no chance moves.<sup>8</sup> We use the following notation:

- $i \in I$ , players,
- $h \in \mathcal{H}_i$ , information sets for player  $i$ ,
- $s_i \in S_i$ , pure strategies for player  $i$ ,
- $S = \prod_{i \in I} S_i$ ,  $S_{-i} = \prod_{j \neq i} S_j$ ,
- $U_i : S \rightarrow \mathbb{R}$ , strategic form payoff function for player  $i$ ,
- $s \in S(h)$ , strategy profiles reaching  $h \in \bigcup_{i \in I} \mathcal{H}_i$ ,
- $\mathcal{H}_i(s_i) = \{h \in \mathcal{H}_i : \exists s_{-i} \in S_{-i} \text{ such that } (s_i, s_{-i}) \in S(h)\}$  collects all information sets owned by player  $i$  which strategy  $s_i \in S_i$  does not prevent from being reached.

---

<sup>7</sup>In this particular game our axioms imply that, conditional on *each* node, the player who owns that node would expect the subgame perfect continuation. But this is not true in general: see the discussion following Corollary 4.5 in Section 4.2.

<sup>8</sup>All the results of this section hold when the set of strategy profiles is a Polish space and the collection of information sets is countable. We can also use this framework to represent incomplete information games without a common prior by modeling Nature as an indifferent player who moves before the “real” players.

By perfect recall, for each player  $i$  and each information set  $h \in \mathcal{H}_i$ ,  $S(h) = S_i(h) \times S_{-i}(h)$ , where  $S_i(h)$  and  $S_{-i}(h)$  are the projections of  $S(h)$  on  $S_i$  and  $S_{-i}$  respectively.

Our notation is consistent with the possibility that some moves are simultaneous and that some information sets may be owned by several players, a possibility which is allowed by some extensive form representations of dynamic games (e.g. Osborne and Rubinstein [16], Chapter 6). For example, in multistage games with observable actions we would have  $\mathcal{H}_i = \mathcal{H}$  for all players  $i$ , where  $\mathcal{H}$  is the set of partial histories of action profiles. In this particular case each  $h \in \mathcal{H}$  represents a common observation by all the players.<sup>9</sup>

## 3.2. Infinite hierarchies of conditional beliefs

### 3.2.1. Conditional Probability Systems

Consider a Polish (complete, separable, metrizable) space  $Y$ . We interpret  $y \in Y$  as an *unobservable* (and payoff irrelevant) parameter representing the conditional beliefs of some players' opponents. The Cartesian product  $S \times Y$  is also Polish (of course, we consider the discrete topology on  $S$  and the product topology on  $S \times Y$ ). For each player  $i$ ,

$$\mathcal{B}(\mathcal{H}_i) = \{B : \exists h \in \mathcal{H}_i, B = S(h) \times Y\}$$

is the collection of observable events concerning  $(s, y) \in S \times Y$ . Let  $\mathcal{A}$  be the Borel  $\sigma$ -algebra on  $S \times Y$ . A *conditional probability system* (or CPS) on  $(S \times Y, \mathcal{A}, \mathcal{H}_i)$  is a map

$$\mu(\cdot|\cdot) : \mathcal{A} \times \mathcal{B}(\mathcal{H}_i) \rightarrow [0, 1]$$

satisfying the following axioms:

**Axiom 1.** For all  $B \in \mathcal{B}(\mathcal{H}_i)$ ,  $\mu(B|B) = 1$ .

**Axiom 2.** For all  $B \in \mathcal{B}(\mathcal{H}_i)$ ,  $\mu(\cdot|B)$  is a probability measure on  $(S \times Y, \mathcal{A})$ .

**Axiom 3.** For all  $A \in \mathcal{A}$ ,  $B, C \in \mathcal{B}(\mathcal{H}_i)$ ,  $A \subset B \subset C \Rightarrow \mu(A|B)\mu(B|C) = \mu(A|C)$ .

The set of probability measures on a measure space  $(Z, \mathcal{A})$  is denoted by  $\Delta(Z)$ ; the set of conditional probability systems on  $(S \times Y, \mathcal{A}, \mathcal{H}_i)$  can be regarded as a subset of  $[\Delta(S \times Y)]^{\mathcal{H}_i}$

---

<sup>9</sup>Battigalli [6] analyzes multistage games with observed actions and incomplete information.

(the set of mappings from  $\mathcal{H}_i$  to  $\Delta(S \times Y)$ ) and it is denoted by  $\Delta^{\mathcal{H}_i}(S \times Y)$ .<sup>10</sup> Accordingly, we often write  $\mu = (\mu(\cdot|S(h) \times Y))_{h \in \mathcal{H}_i} \in \Delta^{\mathcal{H}_i}(S \times Y)$ . The topology on  $S \times Y$  and  $\mathcal{A}$ , the smallest sigma-algebra containing it, are always understood and need not be explicit in our notation. Thus we simply say “conditional probability system (or CPS) on  $(S \times Y, \mathcal{H}_i)$ ...” It is also understood that  $\Delta(S \times Y)$  is endowed with the topology of weak convergence of measures and  $[\Delta(S \times Y)]^{\mathcal{H}_i}$  is endowed with the product topology. Thus  $\Delta(S \times Y)$  and  $[\Delta(S \times Y)]^{\mathcal{H}_i}$  (by countability of  $\mathcal{H}_i$ ) are Polish spaces. Since  $\Delta^{\mathcal{H}_i}(S \times Y)$  is a closed subset of  $[\Delta(S \times Y)]^{\mathcal{H}_i}$ , also  $\Delta^{\mathcal{H}_i}(S \times Y)$  is a Polish space (endowed with the relative topology inherited from  $[\Delta(S \times Y)]^{\mathcal{H}_i}$ ). The set  $X = S \times Y \times \Delta^{\mathcal{H}_i}(S \times Y)$  endowed with the product topology is also a Polish space. We interpret a point  $(s, y, \mu) \in X$  as a state specifying which actions the players would choose at any information set  $(s)$ , the (yet unmodeled) beliefs of  $i$ 's opponents  $(y)$  and the conditional beliefs on  $(S \times Y)$  that player  $i$  would have at each information set  $h \in \mathcal{H}_i$ .

### 3.2.2. Inductive Construction

Here we generalize the inductive construction of a universal type space provided by Brandenburger and Dekel [11] (henceforth BD). For all  $i \in I$  and all nonnegative integers  $n$ ,

$$\begin{aligned} X_i^0 &= S, \\ X_i^{n+1} &= X_i^n \times \prod_{j \neq i} \Delta^{\mathcal{H}_j}(X_j^n). \end{aligned}$$

The set of infinite hierarchies of CPSs for player  $i$  is  $H_i = \prod_{n=0}^{\infty} \Delta^{\mathcal{H}_i}(X_i^n)$ . An infinite hierarchy represents an epistemic type and is therefore typically denoted by  $t_i = (\mu_i^1, \mu_i^2, \dots, \mu_i^n, \dots)$ . Note that for all  $n \geq 0$ ,  $X_i^n$  and  $\Delta^{\mathcal{H}_i}(X_i^n)$  are Polish spaces. It follows that also  $H_i$  and  $\Delta^{\mathcal{H}_i}(S \times \prod_{j \neq i} H_j)$  are Polish spaces. Note also that for all  $n > k \geq 0$ ,  $X_i^n$  can be decomposed as follows:

$$X_i^n = X_i^k \times \prod_{l=k}^{n-1} \prod_{j \neq i} \Delta^{\mathcal{H}_j}(X_j^l).$$

Similarly,  $S \times \prod_{j \neq i} H_j$  is homeomorphic to  $X_i^k \times \prod_{j \neq i} \prod_{l=k}^{\infty} \Delta^{\mathcal{H}_j}(X_j^l)$ . This is not simply a decomposition because we need to swap coordinates. That is, let  $H_{-i}^* = \prod_{n=0}^{\infty} \prod_{j \neq i} \Delta^{\mathcal{H}_j}(X_j^n)$ ... Clearly  $S \times H_{-i}^*$  is homeomorphic to  $S \times \prod_{j \neq i} H_j$  (via coordinate permutation) and for all

---

<sup>10</sup>The definition of a CPS on  $(S, \mathcal{A}, \mathcal{H}_i)$  is obvious.

$k \geq 0$ ,

$$S \times H_{-i}^* = X_i^k \times \prod_{l=k}^{\infty} \prod_{j \neq i} \Delta^{\mathcal{H}_j}(X_j^l).$$

The canonical homeomorphism from  $S \times H_{-i}^*$  to  $S \times \prod_{j \neq i} H_j$  is denoted by  $\varphi$ .

### 3.2.3. Coherent Hierarchies

We have not yet imposed any coherency condition relating beliefs of different orders. Of course, we want to assume that, conditional on every information set, beliefs of different orders assign the same probability to the same event. For all  $h \in \mathcal{H}_i$  and  $n = 1, 2, \dots, \infty$  let  $\mathcal{C}_i^n(h) \subset X_i^n$  denote the cylinder with base  $S(h)$  in  $X_i^n$ , that is

$$\mathcal{C}_i^n(h) = S(h) \times \prod_{k=0}^{n-1} \prod_{j \neq i} \Delta^{\mathcal{H}_j}(X_j^k).$$

Similarly,  $\mathcal{C}_i^\infty(h) = S(h) \times \prod_{j \neq i} H_j$ . For any probability measure  $\nu$  on a product space  $X \times Y$ ,  $\text{mrg}_X \nu \in \Delta(X)$  denotes the marginal measure on  $X$

**Definition 3.1.** *An infinite hierarchy of CPSs  $t_i = (\mu_i^1, \mu_i^2, \dots, \mu_i^n, \dots)$  is coherent if for all  $h \in \mathcal{H}_i$ ,  $n = 1, 2, \dots$ ,*

$$\text{mrg}_{X_i^{n-1}} \mu_i^{n+1}(\cdot | \mathcal{C}_i^n(h)) = \mu_i^n(\cdot | \mathcal{C}_i^{n-1}(h)). \quad (3.1)$$

The set of coherent hierarchies for  $i$  is denoted by  $H_{i,c}$ .

**Proposition 3.2.** *(cf. BD, Proposition 1) For all  $i \in I$  there exists a canonical homeomorphism  $f_i : H_{i,c} \rightarrow \Delta^{\mathcal{H}_i}(S \times \prod_{j \neq i} H_j)$  such that if  $\mu_i = f_i(\mu_i^1, \mu_i^2, \dots, \mu_i^n, \dots)$ , then for all  $h \in \mathcal{H}_i$ ,  $n = 1, 2, \dots$ ,  $E^{n-1} \subset X_i^{n-1}$  (measurable)*

$$\mu_i \left( \varphi \left( E^{n-1} \times \prod_{l=n-1}^{\infty} \prod_{j \neq i} \Delta^{\mathcal{H}_j}(X_j^l) \right) | \mathcal{C}_i^\infty(h) \right) = \mu_i^n(E^{n-1} | \mathcal{C}_i^{n-1}(h)) \quad (3.2)$$

where  $\varphi$  is the canonical homeomorphism from  $S \times H_{-i}^*$  to  $S \times \prod_{j \neq i} H_j$ .

**Proof.** The proof can be adapted from Battigalli [6] which builds on BD (see also Siniscalchi [22]). ■

We let  $f_{i,h}$  denote the  $h$ -coordinate function derived from  $f_i$ , that is, if  $\mu_i = f_i(t_i)$  then  $f_{i,h}(t_i) = \mu_i(\cdot | \mathcal{C}_i^\infty(h))$  is the probability measure on  $S \times \prod_{j \neq i} H_j$  conditional on  $\mathcal{C}_i^\infty(h) = S(h) \times \prod_{j \neq i} H_j$ .

### 3.2.4. Common Certainty of Coherency

Even if  $i$ 's hierarchy of CPSs,  $t_i$ , is coherent, some elements of  $f_i(t_i)$  (i.e. some  $f_{i,h}(t_i)$ ,  $h \in \mathcal{H}_i$ ) may assign positive probability to sets of incoherent hierarchies of some other player  $j$ . We restrict our attention to the case whereby every player  $i$  conditional on every  $\mathcal{C}_i^\infty(h)$  believes that there is common certainty of coherency.

Player  $i$  endowed with coherent hierarchy of CPSs  $t_i$  is *certain* of some (measurable) event  $E \subset S \times \prod_{j \neq i} H_j$  (concerning  $s$  and/or the other players' beliefs) *given*  $h \in \mathcal{H}_i$  if  $f_{i,h}(t_i)(E) = 1$ . Thus we can give the following inductive definition: for all  $i \in I$ :

$$H_{i,c}^1 = H_{i,c},$$

for all  $k \geq 2$ ,

$$H_{i,c}^k = \{t_i \in H_{i,c}^{k-1} : \forall h \in \mathcal{H}_i, f_{i,h}(t_i)(S \times \prod_{j \neq i} H_{j,c}^{k-1}) = 1\},$$

$$T_i = \bigcap_{k \geq 1} H_{i,c}^k.$$

**Proposition 3.3.** (cf. BD, Proposition 2) For all  $i \in I$ , the restriction of  $f_i = (f_{i,h})_{h \in \mathcal{H}_i}$  to  $T_i \subset H_{i,c}$  induces an homeomorphism  $g_i = (g_{i,h})_{h \in \mathcal{H}_i} : T_i \rightarrow \Delta^{\mathcal{H}_i}(S \times \prod_{j \neq i} T_j)$  (defined by  $g_{i,h}(t_i)(E) = f_{i,h}(t_i)(E)$  for all  $h \in \mathcal{H}_i$ ,  $t_i \in T_i$  and measurable subsets  $E \subset S \times \prod_{j \neq i} T_j$ ).

**Proof.** The proof can be adapted from Battigalli [6]. ■

Proposition 3.3 shows that each element  $t_i \in T_i$  corresponds to an epistemic type in the usual sense, except that here a type  $t_i$  is associated to a conditional probability system on  $(S \times \prod_{j \neq i} T_j, \mathcal{H}_i)$  instead of an ordinary probability measure on  $S \times \prod_{j \neq i} T_j$ . In particular, the *tuple*  $(S, (T_i)_{i \in I}, (g_i)_{i \in I})$  can be regarded as an extensive form epistemic model in the sense of Ben Porath [9].

## 4. Extensive Form Rationalizability

In this section we define and characterize extensive form rationalizability (EFR henceforth). EFR is an iterative deletion procedure. Our definition of EFR is very close to the one originally put forward by Pearce [17] and it is indeed equivalent to it for every two-person game. We provide an epistemic characterization of the strategies surviving each step of the procedure as well as a characterization of the rationalizable strategies. To make our formulation simpler and more transparent the following notation will be convenient. A first order CPS for player  $i$  is typically denoted by  $\delta_i$ , that is,  $\delta_i \in \Delta^{\mathcal{H}_i}(S)$ . For every type

$t_i \in T_i$  and  $h \in \mathcal{H}_i$ ,  $\delta_{i,h}(t_i)$  denotes the marginal of  $g_{i,h}(t_i)$  on  $S$ . It is easy to see that  $\delta_i(t_i) \equiv (\delta_{i,h}(t_i))_{h \in \mathcal{H}_i} \in \Delta^{\mathcal{H}_i}(S)$ . The notation used for profiles of types is analogous to the one used for profiles of strategies:  $t \in T = \prod_{i \in I} T_i$ ,  $t_{-i} = T_{-i} = \prod_{j \neq i} T_j$ .

#### 4.1. The Iterative Deletion Procedure

The basic building block of EFR is the notion of *weak sequential rationality*. This is a best response property which applies to plans of actions<sup>11</sup> as well as strategies (see e.g. Reny [19]). We adopt the specific formalization proposed in Battigalli [6] (see his Definition 5.1):

**Definition 4.1.** Fix a first order CPS  $\delta_i \in \Delta^{\mathcal{H}_i}(S)$ . A strategy  $s_i \in S_i$  is a weakly sequential best reply to  $\delta_i$  iff, for every  $h \in \mathcal{H}_i(s_i)$  and every  $s'_i \in S_i(h)$

1.  $\delta_i(\{s_i\} \times S_{-i}(h) | S(h)) = 1$ .
2.  $\sum_{s_{-i} \in S_{-i}} [U_i(s_i, s_{-i}) - U_i(s'_i, s_{-i})] \delta_i(\{(s_i, s_{-i})\} | S(h)) \geq 0$

We refer the interested reader to [6] for details on the features of this definition. Here we simply point out that part 1 essentially means that a rational player is certain of her strategy (hence of her future contingent choices) as long as she knows that she has not deviated from it. Given this, we can derive a “marginal” first order CPS  $\delta_{s_i}$  on  $(S_{-i}, \mathcal{H}_i(s_i))$  and we require that  $s_i$  be a best response to  $\delta_{s_i}(\cdot | S_{-i}(h))$  at each relevant information set  $h \in \mathcal{H}_i(s_i)$ .

Observe also that, by part 1 of the definition, a weakly sequential best reply to a first-order belief is *unique*, if it exists at all. This makes it possible to define a function  $r_i : \Delta^{\mathcal{H}_i}(S) \rightarrow S_i \cup \{\emptyset\}$  assigning to each  $\delta_i \in \Delta^{\mathcal{H}_i}(S)$  the unique weakly sequential best reply  $s_i$  or the symbol  $\emptyset$ , as the case may be. Also, for every type  $t_i \in T_i$ , we let  $\rho_i(t_i) = r_i(\delta_i(t_i))$  denote the “best reply” to the first order beliefs of type  $t_i$ .

We are now ready to define the solution procedure called “extensive form rationalizability.”

**Definition 4.2.** Let  $S^0 = S$ . Assume that  $S^1, \dots, S^n$  have been defined. Then  $s = (s_i)_{i \in I} \in S^{n+1}$  if and only if  $s \in S^n$  and, for each player  $i$ , there exists some  $\delta_i^n \in \Delta^{\mathcal{H}_i}(S)$  such that:

---

<sup>11</sup>Intuitively, a plan of action for player  $i$  is silent about which actions would be taken by  $i$  if  $i$  did not follow that plan. Formally, a *plan of action* is a class of realization-equivalent strategies. In generic extensive games, a plan of action is a strategy of the reduced normal form.

1. For each  $h \in \mathcal{H}_i(s_i)$ ,  $S(h) \cap S^n \neq \emptyset \Rightarrow \delta_i^n(S^n|S(h)) = 1$ .
2.  $s_i \in r_i(\delta_i^n)$ .

For every player  $i$  and positive integer  $n$ , denote by  $S_i^n$  the projection of  $S^n$  on  $S_i$ . A strategy  $s_i$  for player  $i$  is extensive form rationalizable if  $s_i \in \bigcap_{n>0} S_i^n$ .

The preceding definition is very similar to that originally proposed by Pearce [17]; we deviate from this author in two ways.

First, at each step  $n$ , for every candidate strategy  $s_i$  of Player  $i$  and information set  $h \in \mathcal{H}_i(s_i)$ , Pearce considers the set of surviving strategies  $s'_i \in S_i^{n-1}$  which only differ from  $s_i$  at information sets which (weakly) follow  $h$ . He then requires that  $s_i$  be at least as good a reply as any element of this set (for an appropriate conjecture), but does not require that  $s_i$  be compared with any other strategy in  $S_i(h)$ . Our definition is (apparently) stronger, in that we do require (see Definition 4.1) that  $s_i$  be at least as good as any strategy in  $S_i$  which reaches the information set  $h$ . However, this difference is immaterial (see Battigalli [7]).

A major departure from Pearce's definition is that we consider a notion of *correlated* rationalizability, while Pearce assumes that players' conditional beliefs are derived from *product* priors on the set  $\prod_{i \in I} S_i$ . Clearly, this difference, too, is immaterial in two-person games, and it can be shown that the same holds for perfect information games (see Battigalli [7]). Battigalli [5] argues that Pearce does not appropriately model stochastic independence of beliefs in extensive form games and proposes an alternative notion of rationalizability with independence.

Before we move on to the main section of the paper, let us verify that the iterative deletion procedure just defined singles out the profile ((In, T), L) in the Battle of the Sexes with an outside option and the backward induction outcome in the "Centipede." In the first game we have  $S_1^0 = \{(\text{In}, \text{T}), (\text{In}, \text{B}), (\text{Out}, *)\}$  (the choice after Out is clearly irrelevant as far as weak sequential rationality is concerned) and  $S_2^0 = \{\text{L}, \text{R}\}$ . Since (In, B) is strictly dominated, we have  $S_1^1 = \{(\text{In}, \text{T}), (\text{Out}, *)\}$ , while  $S_2^1 = S_2^0$ . For  $n = 2$ , since the only strategy profiles in  $S^1$  which reach the simultaneous-moves subgame are those in which Player 1 plays In and follows with T,  $S_2^2 = \{\text{L}\}$ ;  $S_1^2 = S_1^1$ . But for  $n = 3$  we finally get  $S_1^3 = \{(\text{In}, \text{T})\}$ , and of course  $S_2^3 = S_2^2$ : we have identified the forward induction solution. In the "Centipede" game extensive form rationalizability is closely related to the backward induction procedure: step 1 eliminates (a,a), step 2 eliminates (A,A), step 3 eliminates (a,d) and, finally, step 4 eliminates (A,D).



## 4.2. Epistemic Characterization

In our epistemic model, the set of *states of the world* is  $S \times T = \prod_{i \in I} S_i \times \prod_{i \in I} T_i$ . An *event* is a measurable subset  $E \subset S \times T$ .  $E_{t_i} \subset S \times T_{-i}$  is event  $E$  from the point of view of type  $t_i$ , that is,

$$E_{t_i} = \{(s, t_{-i}) \in S \times T_{-i} : (s, t_i, t_{-i}) \in E\}.$$

$R_i$  is the event “player  $i$  is rational”, that is,

$$R_i = \{(s, t) \in S \times T : s_i = \rho_i(t_i)\}.$$

$R = \bigcap_{i \in I} R_i$  is the event “everyone is rational.”

The following definition introduces the first key ingredient in our axiomatization. We formalize the idea that a player may formulate a conjecture on her opponents’ strategies and types, and may be unwilling to revise it unless, in the course of the game, she receives information which falsifies it.

**Definition 4.3.** *For any event  $E$  and type  $t_i \in T_i$  we say that type  $t_i$  strongly believes  $E$  (believes  $E$  whenever possible) if for all information sets  $h \in \mathcal{H}_i$ ,*

$$E_{t_i} \cap (S(h) \times T_{-i}) \neq \emptyset \Rightarrow g_{i,h}(t_i)(E_{t_i}) = 1.$$

Let  $\beta_i^*(E)$  denote the event that player  $i$  strongly believes  $E$  and let  $\beta^*(E)$  denote the event that everybody strongly believes  $E$ , that is:

- $\beta_i^*(E) := \{(s, t) : \forall h \in \mathcal{H}_i, E_{t_i} \cap (S(h) \times T_{-i}) \neq \emptyset \Rightarrow g_{i,h}(t_i)(E_{t_i}) = 1\},$
- $\beta^*(E) := \bigcap_{i \in I} \beta_i^*(E).$

By inspecting the definition of strong belief, one notices that the event  $E$  itself determines the class of information sets  $h \in \mathcal{H}_i$  where Player  $i$ ’s conjectures are restricted. Alternative belief operators (see the next subsection, or Battigalli [6]) restrict players’ beliefs on a *fixed* family of information sets.

This simple observation has two important consequences. First, for arbitrary events  $E$  and  $F$ , we have  $\beta_i^*(E \cap F) \supseteq \beta_i^*(E) \cap \beta_i^*(F)$ , but *equality need not hold*. Clearly, every state of the world in  $\beta_i^*(E) \cap \beta_i^*(F)$  is such that Player  $i$ ’s belief at any information set  $h \in \mathcal{H}_i$  consistent with  $E \cap F$  (i.e. such that  $(E \cap F)_{t_i} \cap (S(h) \times T_{-i}) \neq \emptyset$ ) assigns probability one

to both  $E$  and  $F$ , hence to  $E \cap F$ : thus, every such state is also an element of  $\beta_i^*(E \cap F)$ . However, the converse need not be true, because there might be an information set  $h \in \mathcal{H}_i$  which is inconsistent with  $F$  but consistent with  $E$  (or vice versa): in this case,  $\beta_i^*(E \cap F)$  places no restrictions on Player  $i$ 's beliefs at any such  $h$ , but clearly  $\beta_i^*(E) \cap \beta_i^*(F)$  does. Thus, a given state of the world may be an element of the former set, but not of the latter. As a result, one must be careful to interpret axioms involving conjunctions of strong belief operators accurately.

Second, note that the argument above implies that, unlike standard epistemic operators, the strong belief operator  $\beta_i^*$  is not monotone (otherwise the inclusion relation  $\beta_i^*(E \cap F) \supseteq \beta_i^*(E) \cap \beta_i^*(F)$  would necessarily hold as an equality).

The second ingredient in our axiomatization is the *best rationalization principle*. The idea (which will be made explicit in Remark 1 below) is that, at each point in the game, players bestow the highest possible degree of strategic sophistication upon their opponents. That is, every player strongly believes that her opponents are rational, *and* she strongly believes that they, too, strongly believe that their opponents are rational, etc.. Thus, if in the course of the game she receives information which contradicts the latter statement, but not the former, she continues to be certain that the former statement is true.

These observations motivate the next few definitions. For any event  $E$ , let

$$\gamma(E) = E \cap \beta^*(E)$$

denote the set of states where  $E$  is true and everybody strongly believes  $E$ . Note that  $\gamma(E)$  is measurable; also, by construction, it is *monotone*. Since operator  $\gamma$  preserves measurability, we can define iterations of  $\gamma$  in the usual way. In particular we obtain the following identities:

$$\gamma^0(E) = E,$$

$$\gamma^1(E) = E \cap \beta^*(E),$$

$$\gamma^2(E) = \gamma[E \cap \beta^*(E)] = E \cap \beta^*(E) \cap \beta^*[E \cap \beta^*(E)],$$

....

It should be clear that the iterated application of the operator  $\gamma$  yields a sequence of events which represent the restrictions discussed above and in the introduction. We can actually be even more explicit:

**Remark 1.** *By inspection of the definitions above*

$$\gamma^n(R) = \bigcap_{i \in I} \left[ R_i \cap \left( \bigcap_{k=0}^{n-1} \beta_i^*(\gamma^k(R)) \right) \right].$$

Therefore  $(s, t) \in \gamma^n(R)$  if and only if, for each player  $i$ ,  $s_i \in \rho_i(t_i)$  and  $t_i$  strongly believes  $\gamma^k(R)$  for all  $k = 0, \dots, n-1$ .

Having disposed of all the preliminaries, we are ready to state our main result:

**Proposition 4.4.** *For every strategy profile  $s \in S$  the following statements hold:*

- (a) *for all  $n \geq 0$ ,  $s \in S^{n+1}$  if and only if there exists a profile of infinite hierarchies of conditional beliefs  $t \in T$  such that  $(s, t) \in \gamma^n(R)$ ;*
- (b)  *$s \in \bigcap_{n=0}^{\infty} S^n$  if and only if there exists a profile of infinite hierarchies of beliefs  $t \in T$  such that  $(s, t) \in \bigcap_{n=0}^{\infty} \gamma^n(R)$ .*

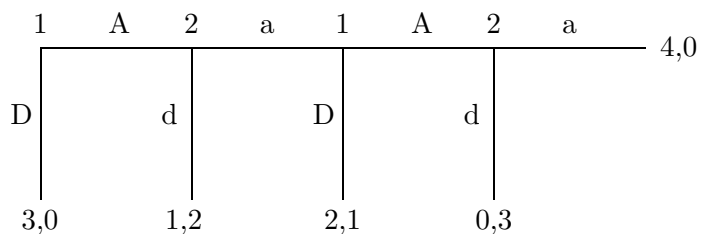
More concisely, we can say that, for every  $n \geq 0$ ,  $S^{n+1}$  is the projection on  $S$  of the event  $\gamma^n(R)$ ; a similar statement holds as  $n \rightarrow \infty$ .

Battigalli [7] shows that, in generic finite games of perfect information, extensive form rationalizability is outcome equivalent to backward induction. Therefore we obtain the following corollary.

**Corollary 4.5.** *Suppose that the given game has perfect information and there are no ties between payoffs at different terminal nodes. Then for every state  $(s, t) \in \bigcap_{n=0}^{\infty} \gamma^n(R)$ , the path induced by  $s$  coincides with the (unique) backward induction path.*

We emphasize that the sufficient conditions stated in Corollary 4.5 are explicit and transparent assumptions about players' dispositions to act  $((s, t) \in R)$  and to form and revise their beliefs when faced with either expected or unexpected evidence  $((s, t) \in \bigcap_{n=1}^{\infty} \gamma^n(R))$ .

Furthermore, our assumptions do not imply that a player at a non rationalizable node would play and/or expect the backward induction continuation. In certain games this is actually *inconsistent* with strong belief in rationality and the best rationalization principle: the following example (Figure 3 in Reny [19]), illustrates the point:



At the first iteration,  $\gamma^0(R) = R$  eliminates (A,D) for Player 1 and (a,a) for Player 2. Then,  $\gamma^1(R) = R \cap \beta^*(R)$  eliminates (A,A) and (d,\*), which yields the backward induction *outcome*. Hence, if Player 2 is reached, he must conclude that it is *not* the case that everybody is rational and there is strong belief in rationality. However, he *can* continue to (strongly) believe that everybody, in particular Player 1, is rational. Of course, this implies that Player 2 should expect Player 1 to move Across at her second node.

The best rationalization principle, embodied in the axiom  $\gamma^2(R) = \gamma^1(R) \cap \beta^*(\gamma^1(R)) = R \cap \beta^*(R) \cap \beta^*(R \cap \beta^*(R))$ , incorporates precisely this restriction. However, backward induction reasoning implies that Player 2, upon being reached, should expect Player 1 to move Down at her next node.

Finally, it should be noted Tan and Werlang's characterization of correlated rationalizability (iterated strict dominance) in normal-form games ([25], Theorem 5.3) may be viewed as a special case of our Proposition 4.4.

In such games, players' conjectures are adequately represented by (single) probability measures, which can be seen as (degenerate) conditional probability systems in which the only conditioning event is the empty history. In such circumstances, strong belief is easily seen to reduce to probability one belief – a monotone epistemic operator.

As a consequence, on one hand our axioms specialize to those proposed by Tan and Werlang; on the other, extensive-form rationalizability coincides with iterated strict dominance in simultaneous games. The characterization result now follows immediately from Proposition 4.4.

### 4.3. Rationalizability and Conditional Common Certainty of Rationality

Let us now restrict our attention to the class of multistage games with *observable actions*. In this case  $\mathcal{H}_i = \mathcal{H}$ ,  $i \in I$ , where  $\mathcal{H}$  is the set of partial (or non terminal) histories. Each  $h \in \mathcal{H}$  represents a common observation and, for any event  $E$ , it makes sense to define the

event “there would be *common certainty* of  $E$  at  $h$ .” Formally,

$$\beta_{i,h}(E) = \{(s, t) : g_{i,h}(t_i)(E_{t_i}) = 1\}$$

is the event “ $i$  would be certain of  $E$  at  $h$ .” This is of course a “standard”, monotone belief operator. Let

$$\begin{aligned} \beta_h(E) &= \bigcap_{i \in I} \beta_{i,h}(E), \\ (\beta_h)^0 &= E \end{aligned}$$

and, for all  $n \geq 0$ ,

$$(\beta_h)^{n+1}(E) = \beta_h((\beta_h)^n(E)).$$

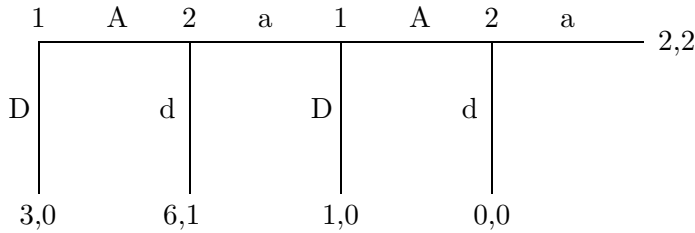
Then event “there would be common certainty of  $E$  at  $h$ ” is

$$c\beta_h(E) = \bigcap_{n > 0} (\beta_h)^n(E).$$

We can then state a simple relationship between extensive form rationalizability and common certainty of rationality. Say that partial history  $h$  is *consistent with rationality and common certainty of rationality* if  $(R \cap c\beta_h(R)) \cap (S(h) \times T) \neq \emptyset$ .

**Proposition 4.6.** *For every multistage game with observable actions, if partial history  $h \in \mathcal{H}$  is induced by some profile of extensive form rationalizable strategies (i.e.  $S^\infty \cap S(h) \neq \emptyset$ ) then  $h$  is consistent with rationality and common certainty of rationality.*

Note that the proposition provides only a sufficient condition. There are games with histories consistent with common certainty of rationality and yet unreachable by profiles of extensive form rationalizable strategies. The following example illustrates this point.<sup>12</sup>



<sup>12</sup>Reny [18] provides a similar example, although his discussion does not employ a formal epistemic model.

Consider the following epistemic model (probabilities which are not relevant to the argument are omitted for simplicity):

	Pair	$p_1(\emptyset)$	$p_1(Aa)$		Pair	$p_2(A)$	$p_2(Aa), p_2(AaA)$
1	$((D,*), t_{11})$	0,1,0	—	1	$((d,*), t_{12})$	0,0,1	—
2	$((A,A), t_{21})$	$\frac{1}{2}, \frac{1}{2}, 0$	0,1,0	2	$((a,a), t_{22})$	0,1,0	0,1,0
3	$((A,D)), t_{31})$	$\frac{1}{2}, \frac{1}{6}, \frac{1}{3}$	$0, \frac{1}{3}, \frac{2}{3}$	3	$((a,d), t_{32})$	—	—

It is easy to see that extensive-form rationalizability requires Player 1 to choose D at her first node (if Player 2's first node were reached, then the game would be expected to continue with a, A and a). Thus, Player 1's second node (following the history Aa) cannot be reached.

Yet, in state (2,2) there is common certainty of rationality at Player 1's second node, as is apparent from the conditional probabilities  $p_1(Aa)$  and  $p_2(Aa)$ .

To verify directly that state (2,2) is indeed inconsistent with our axioms, notice first that our axioms eliminate all states  $(i, 3)$  because  $(a,d)$  is (sequentially) irrational for Player 2, regardless of his conjecture. Then, in successive iterations, this leads to the elimination of states involving types  $t_{31}$ ,  $t_{12}$ , and finally  $t_{21}$ .

## 5. Discussion and Conclusions

We have provided an epistemic characterization of extensive-form rationalizability which highlights its relationship with common certainty of rationality and forward-induction reasoning. As a by-product, we have also obtained sufficient conditions for the backward induction outcome to obtain in generic  $N$ -person games with perfect information.

The strong belief operator plays a central role in our analysis (but see also Siniscalchi [22] and Stalnaker [24].) We have argued that the underlying idea of “belief whenever possible” is quite natural in extensive-form analysis, and is essentially indispensable if one wishes to discuss forward induction concepts such as the best rationalization principle.

Although our analysis is restricted to finite games with complete information, its extensions to a more general class of games is entirely straightforward. Battigalli [6] already provides an epistemic analysis of dynamic games of incomplete information. Our construction of the universal, extensive form type space can be carried out under fairly general

assumptions (i.e.  $S_i$  Polish and  $\mathcal{H}_i$  countable,  $i \in I$ ). Given this, we conjecture that our main result can be proved under standard compactness-continuity assumptions.

According to the notion of rationalizability discussed here, a player may have correlated beliefs about his opponents. Moreover, the best rationalization principle, as axiomatized here, also reflects a notion of “correlated” belief revision: for instance, if a player observes an irrational move by one of her opponents, she is *not* required to maintain her belief in the rationality of her other opponents.

This is perfectly consistent with a noncooperative approach (e.g. Stalnaker [23] and [24]), especially outside the realm of equilibrium analysis. Also, our results imply that neither aspect is actually crucial in order to obtain the backward induction outcome in generic perfect information games. Finally, we have already pointed out that our definition is actually equivalent to that originally proposed by Pearce [17] in general games with perfect information.

However, at the heart the best rationalization principle is the assumption that players tend to attribute the *highest possible degree of strategic sophistication* to their opponents. Hence, a notion of stochastic independence and independent revision of beliefs about distinct opponents seems to be called for, perhaps even as a matter of consistency.

In order to focus on the somewhat more basic notions of strong belief and best rationalization, we have relegated these issues to a companion paper [8]. There we characterize a solution concept proposed by Battigalli [5] and (modulo some technical differences) Reny [19], which incorporates both forward induction ideas and independent beliefs.

Another natural direction for further research is to examine *equilibrium* concepts which incorporate notions of forward induction, such as Cho and Kreps’s [12] well-known “intuitive criterion”.

## References

- [1] ASHEIM, G.B. and M. DUFWENBERG (1996): “Admissibility and Common Knowledge,” mimeo, Department of Economics, University of Oslo.
- [2] AUMANN, R.J. (1995): “Backward Induction and Common Knowledge of Rationality,” *Games and Economic Behavior*, **8**, 6-19.

- [3] AUMANN, R.J. (1996): "Reply to Binmore," *Games and Economic Behavior*, **17**, 138-146.
- [4] AUMANN, R.J. (1996): "Deriving Backward Induction in the Centipede Game without Assuming Rationality at Unreached Vertices," mimeo, The Hebrew University, Jerusalem.
- [5] BATTIGALLI, P. (1996): "Strategic Rationality Orderings and the Best Rationalization Principle," *Games and Economic Behavior*, **13**, 178-200.
- [6] BATTIGALLI, P. (1996): "Hierarchies of Conditional Beliefs and Interactive Epistemology in Dynamic Games," mimeo, Princeton University and IGIER, Milan.
- [7] BATTIGALLI, P. (1997): "On Rationalizability in Extensive Games," *Journal of Economic Theory*, **74**, 40-60.
- [8] BATTIGALLI, P. and M. SINISCALCHI (1997): "Interactive Beliefs, Epistemic Independence and Explicability," mimeo, Princeton University and Stanford University.
- [9] BEN PORATH, E. (1997): "Rationality, Nash Equilibrium and Backwards Induction in Perfect Information Games," *Review of Economic Studies*, **64**, 23-46.
- [10] BERNHEIM, D. (1984): "Rationalizable Strategic Behavior," *Econometrica*, **52**, 1002-1028.
- [11] BRANDENBURGER, A. and E. DEKEL (1993): "Hierarchies of Beliefs and Common Knowledge," *Journal of Economic Theory*, **59**, 189-198.
- [12] CHO, I.-K. and D. KREPS (1987): "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, **102**, 179-221.
- [13] DEKEL, E. and F. GUL (1997): "Rationality and Knowledge in Game Theory," in *Advances in Economics and Econometrics* (D. Kreps and K. Wallis, Eds.). Cambridge UK: Cambridge University Press.
- [14] FAGIN, R., J. HALPERN, Y. MOSES and M. VARDI (1995): *Reasoning about Knowledge*. Cambridge MA: M.I.T. Press.



- [15] MYERSON, R. (1986): “Multistage Games with Communication,” *Econometrica*, **54**, 323-358.
- [16] OSBORNE, M. and A. RUBINSTEIN (1994): *A Course in Game Theory*. Cambridge MA: MIT Press.
- [17] PEARCE, D. (1984): “Rationalizable Strategic Behavior and the Problem of Perfection,” *Econometrica*, **52**, 1029-1050.
- [18] RENY, P. (1985): “Rationality, Common Knowledge and the Theory of Games,” mimeo, Department of Economics, Princeton University.
- [19] RENY, P. (1992): “Backward Induction, Normal Form Perfection and Explicable Equilibria,” *Econometrica*, **60**, 626-649.
- [20] RËNYI, A. (1956): “On Conditional Probability Spaces Generated by a Conditionally Ordered Set of Measures,” *Theory of Probability and Its Applications*, **1**, 61-71.
- [21] SAMET, D. (1996): “Hypothetical Knowledge and Games with Perfect Information,” *Games and Economic Behavior*, **17**, 230-251.
- [22] SINISCALCHI, M. (1997): “Knowledge, Rationality and Heuristics in Extensive Games. Part I: Framework and Non-Equilibrium Analysis,” mimeo, Stanford University.
- [23] STALNAKER, R. (1996): “Knowledge, Belief and Counterfactual Reasoning in Games,” *Economics and Philosophy*, **12**, 133-163.
- [24] STALNAKER, R. (1996): “Belief Revision in Games: Forward and Backward Induction,” mimeo, MIT.
- [25] TAN, T. and S. WERLANG (1988): “The Bayesian Foundation of Solution Concepts of Games,” *Journal of Economic Theory*, **45**, 370-391.

## Appendix

### Proof of Proposition 4.4

We need the following preliminary result:

**Lemma 5.1.** Fix, for every  $j \neq i$ , a collection  $\{t_j(s_j)\}_{s_j \in S_j}$  of types in  $T_j$ . Also, fix a CPS  $\delta_i \in \Delta^{\mathcal{H}_i}(S)$ . Then there exists a unique  $t_i \in T_i$  such that, for each  $h \in \mathcal{H}_i$ ,  $g_{i,h}(t_i)$  has finite support and

$$g_{i,h}(t_i)(s, (t_j(s_j))_{j \neq i}) = \delta_i(s|S(h))$$

for all  $s = (s_j)_{j \in I} \in S$ .

**Proof.** We follow [22], proof of Lemma 7, with the required adjustments. Define a candidate CPS  $\mu_i \in \Delta^{\mathcal{H}_i}(S \times T_{-i})$  by setting

$$\mu_i(s, (t_j(s_j))_{j \neq i} | S(h) \times T_{-i}) = \delta_i(s|S(h))$$

for every  $h \in \mathcal{H}_i$ , and extending the assignments by additivity. Axioms 1 and 2 follow immediately from the observation that the maps  $s_j \mapsto t_j(s_j)$  collectively yield an embedding of  $\bigcup_{h \in \mathcal{H}_i} \text{supp}[\delta_i(\cdot|S(h))]$  (a finite set) in  $S \times T_{-i}$ , so that, for every  $h \in \mathcal{H}_i$ ,  $\mu_i(\cdot|S(h) \times T_{-i})$  is indeed a probability measure on  $S \times T_{-i}$ . By the same argument,  $\mu_i$  must also satisfy Axiom 3, i.e. it must be a CPS; of course, each  $\mu_i(\cdot|S(h) \times T_{-i})$  has finite support by construction. Now invoke the fact that  $g_i$  is a homeomorphism to obtain the required type  $t_i$ . ■

We are now ready to prove the main result.<sup>13</sup>

Part (a) of the proposition is implied by the following statement:

For all  $k \geq 0$ ,

(a1) there are  $2|I|$  functions  $\delta_i^k : S_i \rightarrow \Delta^{\mathcal{H}_i}(S)$  and  $t_i^k : S_i \rightarrow T_i$ ,  $i \in I$ , such that, for all  $s \in S$  and  $i \in I$ ,  $\delta_i^k(s_i) = \delta_i(t_i^k(s_i))$ <sup>14</sup> and

- if  $k \geq 1$ , then for all  $m = 0, \dots, k-1$ , if  $s \in S^m \setminus S^{m+1}$  then  $\delta_i^m(s_i) = \delta_i^k(s_i)$ ,  $t_i^m(s_i) = t_i^k(s_i)$ ,
- if  $s \in S^{k+1}$ , then for all  $h \in \mathcal{H}_i(s_i)$ , all  $m = 0, \dots, k$ , if  $S^m \cap S(h) \neq \emptyset$  then  $\delta_i^k(s_i)(S^m|S(h)) = 1$ ,
- if  $s \in S^{k+1}$  then  $(s, (t_i^k(s_i))_{i \in I}) \in \gamma^k(R)$ ;

(a2) for all states  $(s, t)$ , if  $(s, t) \in \gamma^k(R)$  then  $s \in S^{k+1}$ .

---

<sup>13</sup>For alternative proofs of similar results see [6] and [22].

<sup>14</sup>Recall that  $\delta_i(t_i)$  is the first order CPS of epistemic type  $t_i$ .

( $k = 0$ ), (a1.0). For each  $i$ , fix arbitrary functions  $\delta_i^* : S_i \rightarrow \Delta^{\mathcal{H}_i}(S)$ ,  $t_i^* : S_i \rightarrow T_i$  such that, for all  $s_i$ ,  $\delta_i^*(s_i) = \delta_i(t_i^*(s_i))$  (this is possible by Lemma 5.1). Let  $(\delta_i^0(\cdot), t_i^0(\cdot))$  coincide with  $(\delta_i^*(\cdot), t_i^*(\cdot))$  on  $S_i \setminus S_i^1$ . For each  $i \in I$ , and each  $s_i \in S_i^1$ , let  $\delta_i^0(s_i)$  be a first order CPS rationalizing  $s_i$ , i.e.  $\delta_i \in r_i^{-1}(s_i)$ . Then we can use Lemma 5.1 to define  $t_i = t_i^0(s_i)$  as the (unique)  $t_i$  satisfying:

$$\forall s' \in S, g_{i,h}(t_i) ((s', (t_j^*(s'_j))_{j \neq i})) = \delta_i^0(s_i)(s'|S(h)).$$

By construction,  $\delta_i^0(s_i) = \delta_i(t_i^0(s_i))$  and  $s_i = \rho_i(t_i)$ . Thus, for all  $s \in S^1$ ,  $(s, (t_i^0(s_i))_{i \in I}) \in R = \gamma^0(R)$ . The other statements in (a1) hold trivially.

(a2.0). Suppose that  $(s, t) \in R = \gamma^0(R)$ . Then, for each  $i \in I$ ,  $s_i$  is a weakly sequential best response to  $\delta_i(t_i)$  and thus  $s \in S^1$ .

( $k = n$ ). Suppose that the statement above holds for all  $k = 0, 1, \dots, n-1$ .

(a1.n). For each  $i \in I$  and each  $s_i \in S_i \setminus S_i^{n+1}$ , let  $\delta_i^n(s_i) = \delta_i^{n-1}(s_i)$  and  $t_i^n(s_i) = t_i^{n-1}(s_i)$ . Clearly the first claim in (a1) holds for the functions  $\delta_i^n(\cdot)$  and  $t_i^n(\cdot)$ , independently of how we complete their specification on  $S_i^{n+1}$ .

Let  $s \in S^{n+1}$ . For each  $i \in I$ , define  $\delta_i^n(s_i)$  and  $t_i^n(s_i)$  as follows. Take a first order belief  $\delta_i^n$  satisfying the conditions of Definition 3.2 with respect to  $s_i$ , that is,  $\delta_i^n \in r_i^{-1}(s_i)$  and for all  $h \in \mathcal{H}_i(s_i)$ , if  $S(h) \cap S^n \neq \emptyset$  then  $\delta_i^n(S^n|S(h)) = 1$ . Now define  $\delta_i^n(s_i) \in [\Delta(S)]^{\mathcal{H}_i}$  as follows: for all  $h \in \mathcal{H}_i$  and all  $s' \in S$ ,

$$\delta_i^n(s_i)(s'|S(h)) = \begin{cases} \delta_i^n(s'|S(h)), & \text{if } h \in \mathcal{H}_i(s_i) \text{ and } S^n \cap S(h) \neq \emptyset \\ \delta_i^{n-1}(s_i)(s'|S(h)), & \text{otherwise} \end{cases}.$$

It can be easily checked that  $\delta_i^n(s_i)$  is a CPS, i.e.  $\delta_i^n(s_i) \in \Delta^{\mathcal{H}_i}(S)$ . Furthermore  $s_i = r_i(\delta_i^n(s_i))$ . To see this, note that the inductive hypothesis implies  $s_i = \rho_i(t_i^{n-1}(s_i))$ . Since  $\delta_i^n(s_i)$  is a CPS given by a combination of  $\delta_i^n$  and  $\delta_i^{n-1}(s_i) = \delta_i(t_i^{n-1}(s_i))$  the weak sequential rationality property is satisfied by construction. By Lemma 5.1 we can choose  $t_i = t_i^n(s_i)$  as the (unique) type  $t_i$  satisfying:

$$\forall s' \in S, g_{i,h}(t_i) ((s', (t_j^{n-1}(s'_j))_{j \neq i})) = \delta_i^n(s_i)(s'|S(h)).$$

By construction,  $s_i = \rho_i(t_i^n(s_i))$ ; in light of Remark 1, this provides the first step in the proof of the third claim of (a1.n).

As for the second claim, suppose that  $h \in \mathcal{H}_i(s_i)$  and  $S^n \cap S(h) \neq \emptyset$ . Since the sequence  $\{S^m\}_{m \geq 0}$  is weakly decreasing, it is also the case that for all  $m = 0, \dots, n-1$ ,  $S^m \cap S(h) \neq \emptyset$  and by construction

$$\delta_i^n(s_i)(S^m|S(h)) = \delta_i^n(S^n|S(h)) = 1.$$

Now suppose that, for some  $m = 0, \dots, n-1$ ,  $S^m \cap S(h) \neq \emptyset$  and  $S^n \cap S(h) = \emptyset$ . Then the construction of  $\delta_i^n(s_i)$  and the inductive hypothesis yield

$$\delta_i^n(s_i)(S^m|S(h)) = \delta_i^{n-1}(s_i)(S^m|S(h)) = 1.$$

and this part of the proof is complete.

To prove that  $(s, (t_i^n(s_i)_{i \in I}) \in \gamma^n(R)$  it remains to be shown that for all  $k = 0, \dots, n-1$ ,  $i \in I$  and  $h \in \mathcal{H}_i$ , if  $t_i = t_i^n(s_i)$  and  $[\gamma^k(R)]_{t_i} \cap (S(h) \times T_{-i}) \neq \emptyset$  then  $g_{i,h}(t_i) ([\gamma^k(R)]_{t_i}) = 1$  (see Remark 1). Thus fix  $k \in \{0, \dots, n-1\}$ ,  $i$  and  $h \in \mathcal{H}_i$ ; let  $t_i$  satisfy the foregoing assumption. First note that this implies that  $h \in \mathcal{H}_i(s_i)$ . In fact  $[\gamma^k(R)]_{t_i} \subset [R_i]_{t_i}$ ; since  $s_i = \rho_i(t_i)$ ,  $[R_i]_{t_i} = \{(s', t'_{-i}) : s'_i = s_i = \rho_i(t_i)\}$ . By assumption,  $[R_i]_{t_i} \cap (S(h) \times T_{-i}) \neq \emptyset$ . Thus there is some  $s'_{-i}$  such that  $(s_i, s'_{-i}) \in S(h)$ ; that is,  $h \in \mathcal{H}_i(s_i)$ . By the inductive hypothesis  $S^{k+1}$  is the projection of  $\gamma^k(R)$  on  $S$ . Thus,  $[\gamma^k(R)]_{t_i} \cap (S(h) \times \prod_{j \neq i} T_j) \neq \emptyset$  implies  $S^{k+1} \cap S(h) \neq \emptyset$ . We have just proved that in this case  $\delta_i^{k+1}(s_i)(S^{k+1}|S(h)) = 1$ . Furthermore the inductive hypothesis implies that

$$\forall m \in \{k, \dots, n-2\}, \forall s' \in S^{m+1} \setminus S^{m+2},$$

$$(s', (t_j^{n-1}(s'_j))_{j \in I}) = (s', (t_j^m(s'_j))_{j \in I}) \in \gamma^m(R) \subset \gamma^k(R)$$

and

$$\forall s' \in S^n, (s', (t_j^{n-1}(s'_j))_{j \in I}) \in \gamma^{n-1}(R) \subset \gamma^k(R).$$

Therefore

$$\forall s' \in S^{k+1}, (s', (t_j^{n-1}(s'_j))_{j \in I}) \in \gamma^k(R).$$

These facts and the definition of  $g_{i,h}(t_i) = g_{i,h}(t_i^n(s_i))$  yield

$$\begin{aligned} g_{i,h}(t_i) ([\gamma^k(R)]_{t_i}) &= g_{i,h}(t_i) (\{(s', t'_{-i}) : s'_i = s_i, s' \in S^{k+1}, t'_{-i} = (t_j^{n-1}(s'_j))_{j \neq i}\}) \\ &= \sum_{s' \in S^{k+1}} \delta_i^n(s'|S(h)) = 1 \end{aligned}$$

as desired.

(a2.n). Suppose that  $(s, t) \in \gamma^n(R)$ . Since  $\gamma^n(R) \subset \gamma^{n-1}(R)$ , the inductive hypothesis implies that  $s \in S^n$ . For each  $i$ , let  $\delta_i^n = \delta_i(t_i)$  be the first order conditional beliefs of type  $t_i$ . Since  $\gamma^n(R) \subset R$ ,  $s_i$  is a weakly sequential best response to  $\delta_i^n$ . In order to prove that  $s_i \in S_i^{n+1}$  it is sufficient to check that  $\delta_i^n$  satisfies condition 1 of Definition 3.2. Consider an information set  $h \in \mathcal{H}_i(s_i)$  such that  $S^n \cap S(h) \neq \emptyset$ . By the inductive hypothesis  $S^n$  is the projection on  $S$  of  $\gamma^{n-1}(R)$ . Therefore

$$\gamma^{n-1}(R) \cap (S(h) \times \prod_{j \in I} T_j) \neq \emptyset.$$

Since  $(s, (t_j)_{j \in I}) \in \gamma^n(R) \subset \gamma^{n-1}(R)$ , this implies

$$[\gamma^{n-1}(R)]_{t_i} \cap (S(h) \times \prod_{j \neq i} T_j) \neq \emptyset.$$

Since  $t_i$  strongly believes  $[\gamma^{n-1}(R)]$ , it follows that

$$g_{i,h}(t_i) ([\gamma^{n-1}(R)]_{t_i}) = 1.$$

Taking again into account that  $S^n$  is the projection of  $\gamma^{n-1}(R)$  and that  $\delta_i^n = \delta_i(t_i)$  we obtain  $\delta_i^n(S^n | S(h)) = 1$  as desired. This concludes the proof of part (a).

Part (b) is more straightforward. Note first that  $\bigcap_{n \geq 0} \gamma^n(R)$  is nonempty (this is not trivial because  $T_i$  is uncountable). The key observation is that the set of all probability measures on a compact Polish space is also compact and Polish. In our simple setting,  $S$  is trivially compact, so applying this fact iteratively, using Tychonoff's theorem and noticing that each  $T_i$  is a closed subset of  $H_i$ , one readily sees that  $T_i$  is compact. Hence, since the nonempty closed sets  $\gamma^n(R)$  form a family with the finite intersection property (because they are nested), they have a nonempty infinite intersection.

Now suppose  $(s, t) \in \bigcap_{n \geq 0} \gamma^n(R)$ . Since, by part (a), each  $S^{n+1}$  is the projection on  $S$  of  $\gamma^n(R)$ , we conclude that  $s \in S^{n+1}$  for every  $n$ , so  $s \in \bigcap_{n \geq 0} S^n$  (recall  $S^0 = S$ ).

Finally, let  $N$  be the smallest integer such that  $S^N = \bigcap_{n \geq 0} S^n$  (which must exist because  $S$  is finite). Pick any  $s \in S^N$  and consider the sequence of sets  $M(m, s) = \gamma^{N-1+m}(R) \cap (\{s\} \times T)$ . By part (a), each set  $M(m, s)$  is nonempty and closed; also, the sequence of sets  $M(m, s)$  is decreasing, and hence has the finite intersection property. Then any strategy-type profile in  $\bigcap_{m \geq 0} M(m, s) \neq \emptyset$  has the required properties. ■

### Proof of Proposition 4.6

The proof of Proposition 4.6 relies on the following lemmata. For any event  $E$  let  $\gamma^\infty(E) = \bigcap_{n \geq 0} \gamma^n(E)$ .

**Lemma 5.2.** *For every event  $E$ ,  $\gamma(\gamma^\infty(E)) = \gamma^\infty(E)$ .*

**Proof.** By definition, for every event  $F$ ,  $\gamma(F) = F \cap \beta^*(F) \subset F$ . Thus we only have to show that  $\gamma^\infty(E) \subset \gamma(\gamma^\infty(E))$ . Suppose that  $(s, t) \in \gamma^\infty(E)$ . We must prove that, for all players  $i$ ,  $(s, t) \in \beta_i^*(\gamma^\infty(E))$ , that is, each  $t_i$  in  $t$  strongly believes  $\gamma^\infty(E)$ . Fix  $i$  and  $h \in \mathcal{H}_i$  arbitrarily. Assume that  $[\gamma^\infty(E)]_{t_i} \cap (S(h) \times T) \neq \emptyset$ . Since for all  $n$ ,  $[\gamma^\infty(E)]_{t_i} \subset [\gamma^n(E)]_{t_i}$ , it follows that

$$[\gamma^n(E)]_{t_i} \cap (S(h) \times T) \neq \emptyset. \quad (5.1)$$

Since  $(s, t) \in \gamma^\infty(E) \subset \gamma^{n+1}(E)$ ,  $t_i$  strongly believes  $\gamma^n(E)$  and thus 5.1 implies

$$g_{i,h}(t_i) ([\gamma^n(E)]_{t_i}) = 1. \quad (5.2)$$

As  $\{\gamma^n(E)\}_{n \geq 0}$  is a decreasing sequence of events converging to  $\gamma^\infty(E)$  and the probability measure  $g_{i,h}(t_i)$  is continuous

$$g_{i,h}(t_i) ([\gamma^\infty(E)]_{t_i}) = \lim_{n \rightarrow \infty} g_{i,h}(t_i) ([\gamma^n(E)]_{t_i}) = 1.$$

Therefore  $t_i$  strongly believes  $\gamma^\infty(E)$ . ■

**Lemma 5.3.** *For every multistage game with observed actions, every partial history  $h \in \mathcal{H}$  and every event  $E$ ,*

$$\gamma^\infty(E) \cap (S(h) \times T) \subset \bigcap_{n \geq 0} (\beta_h)^n(E).$$

**Proof.** It is true by definition that  $\gamma^\infty(E) \cap (S(h) \times T) \subset E = (\beta_h)^0(E)$ . Assume that  $\gamma^\infty(E) \cap (S(h) \times T) \not\subset (\beta_h)^n(E)$ . Consider any state  $(s, t) \in \gamma^\infty(E) \cap (S(h) \times T)$ . We must show that, for all  $i$ ,  $g_{i,h}(t_i) ([(\beta_h)^n(E)]_{t_i}) = 1$ . By assumption

$$[\gamma^\infty(E)]_{t_i} \cap (S(h) \times T) \neq \emptyset.$$

By Lemma 5.2  $(s, t) \in \gamma[\gamma^\infty(E)]$ . It follows that

$$g_{i,h}(t_i) ([\gamma^\infty(E)]_{t_i}) = 1.$$

By the inductive hypothesis

$$[\gamma^\infty(E) \cap (S(h) \times T)]_{t_i} \subset [(\beta_h)^n(E)]_{t_i}.$$

Since, by Axiom 1 in the definition of a CPS,  $g_{i,h}(t_i)(S(h) \times T_{-i}) = 1$ , we conclude that

$$g_{i,h}(t_i) ([(\beta_h)^n(E)]_{t_i}) = 1.$$

■

**Proof of Proposition 4.6.** Suppose that  $S^\infty \cap S(h) \neq \emptyset$ . By Proposition 4.4 there is some state  $(s, t) \in \gamma^\infty(R) \cap (S(h) \times T)$ . Then Lemma 5.3 implies that

$$(s, t) \in \bigcap_{n \geq 0} (\beta_h)^n(R) = R \cap c\beta_h(R).$$

■