Research in Economics (1999) **53**, 247–273
Article No. reec.1999.0201
Available online at http://www.idealibrary.com on IDEAL®

Interactive beliefs, epistemic independence and strong rationalizability†

PIERPAOLO BATTIGALLI‡ and MARCIANO SINISCALCHI§

‡*Princeton University and European University Institute*
§*Princeton University*

Summary

We use a universal, extensive form interactive beliefs system to provide an epistemic characterization of a weak and a strong notion of rationalizability with independent beliefs. The weak solution concept is equivalent to backward induction in generic perfect information games where no player moves more than once in any play. The strong solution concept is related to explicability (Reny, 1992) and is outcome-equivalent to backward induction in generic games of perfect information. © 1999 Academic Press

J.E.L. Classification: C72, D82.
Keywords: Interactive epistemology, forward induction, rationalizability, independence, explicability.

1. Introduction

Extensive-form rationalizability (Pearce, 1984; Battigalli, 1996, 1997) attempts to capture the implications of rationality and common certainty of rationality in extensive games. It incorporates a powerful, yet quite natural notion of forward induction, known as the *best rationalization principle*: the idea that, when faced with unexpected events, players attempt to explain ("rationalize") what has transpired in a manner which is consistent with the highest possible degree of strategic sophistication of their opponents.

Battigalli and Siniscalchi (1997) formalize the best rationalization principle in the framework of the extensive-form epistemic

† We would like to thank Eddie Dekel for helpful discussion. All remaining errors are our own.
‡ E-mail: battigal@iue.it
§ E-mail: marciano@princeton.edu

1090–9443/99/030247 + 27 $30.00/0

model developed in Battigalli and Siniscalchi (1998), and show that, together with the assumption that players choose (weakly) sequentially rational strategies, it completely characterizes extensive-form rationalizability.

In their setup, a player's beliefs about her opponents's strategies and epistemic types are represented by conditional probability systems (see Rênyi, 1956; Myerson, 1986). For games with more than two players, Battigalli and Siniscalchi (1997) allow conditional beliefs about the opponents to exhibit *correlation*. Consequently, the solution concept they characterize is more accurately referred to as "correlated extensive-form rationalizability".

This paper focuses on the interplay between independence and rationalizability in extensive games. First, we propose a notion of *epistemic independence*, and show that (i) rationality, (ii) epistemic independence and (iii) common certainty of rationality and epistemic independence at the beginning of the game, completely characterize *weak rationalizability*, a refinement of a solution concept studied by Ben Porath (1997).†

Second, we formalize a notion of *independent best rationalization* and show that, along with epistemic independence and rationality, it completely characterizes *strong rationalizability*, a solution concept first proposed by Battigalli (1996). Interestingly enough, the algorithmic definition of strong rationalizability was motivated by examples in which Pearce's original procedure failed to capture certain "intuitive" implications of the independence assumption. Strong rationalizability is also related to Reny's *explicability* (see Reny, 1992).

Our notion of epistemic independence, which we adapt from Battigalli (1996), formalizes the idea that players should only revise their beliefs about a particular opponent when they receive information about him or her. Care must be taken in defining the relevant product structure on the state space: Subsection 4.1 discusses the details and the relationship with other notions of independence.

In order to obtain a sound notion of extensive form rationalizability with independent beliefs, epistemic independence has to be combined with a modified version of the best rationalization principle. In fact, Battigalli and Siniscalchi (1997) characterize a "collective" version of this principle which requires a player's revised beliefs to be consistent with the highest degree of strategic sophistication which can be jointly attributed to *all* opponents. It only prescribes a *common lower bound on strategic sophistication*; if an unexpected occurrence proves that some opponents are characterized by only "average" strategic sophistication, but does not falsify the assumption that the remaining opponents are "highly"

† Ben Porath (1997) allows for correlated beliefs.

sophisticated, the observer is assumed to attribute *at least* an "average" degree of sophistication to *all* opponents (as opposed to e.g. "low" sophistication), but is *not required* to differentiate among the two groups. In particular, "hard" information concerning the first group of opponents only may be taken to signal that the remaining players, too, are only endowed with "average" strategic sophistication.

But if players' beliefs about their opponents satisfy the epistemic independence property, it is more natural to assume that, when they rationalize observed behavior, players process information about each one of their opponents separately. Thus, independent best rationalization naturally complements epistemic independence.

The axioms we propose may be informally stated as follows. As in Battigalli and Siniscalchi (1997), for any event E, we say that a player *strongly believes* that E if she is certain that E is true conditional on any information set which is not inconsistent with E.

(0S) For every player i,

  ($0S.i$) Player i is (weakly) sequentially rational and has independent conjectures;

(1S) For every player i,

  ($1S.i$) For every opponent $j \neq i$, player i strongly believes that ($0S.j$); ...

(kS) For every player i,

  (kS.i) For every opponent $j \neq i$, player i strongly believes that ($0S.j$) & ($1S.j$) & ... & (($k-1$)S.j); ...

We remark that our formal notion of "strong belief" is slightly different from the one appearing in Battigalli and Siniscalchi (1997), although the basic intuition is the same. Also, our axioms are necessarily "richer" than those proposed in the aforementioned paper. Consequently, our epistemic model (developed in Section 3) must also be richer; specifically, we must allow players to form conjectures conditional on a wider class of hypotheses.

Analogously to Battigalli and Siniscalchi (1997), our results may also be related to *backward induction*. For generic perfect information games in which each player never moves twice in any realization path, weak rationalizability selects the (unique) backward induction strategy profile. Similarly, for arbitrary games with perfect information and generic payoffs, strong rationalizability is outcome-equivalent to backward induction.† Therefore, our characterization results provide alternative sufficient epistemic conditions for the backward induction outcome.

† This does not follow from results in Battigalli and Siniscalchi (1997), because, strictly speaking, strong rationalizability is not "stronger" than correlated rationalizability—only "different". See Battigalli (1996) for additional discussion.

This paper builds on Battigalli and Siniscalchi (1997, 1998). Our extensive form epistemic model can be regarded as a generalization of the model used by Ben Porath (1997) to characterize common certainty of rationality at the beginning of a perfect information game. Stalnaker (1996*a*,*b*) considers 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 Battigalli & Siniscalchi, 1997, 1998). Stalnaker (1996*b*) puts forward a notion of "robust belief" which corresponds to "strong belief" as defined in Battigalli and Siniscalchi (1997) and briefly discusses the relation between robust belief in rationality and forward induction. Also, he proposes a notion of "epistemic independence" which is closely related to the one developed in Battigalli (1996). Aumann (1995, 1996, 1998) and Samet (1996) use different epistemic models and provide a different set of sufficient conditions for the backward induction outcome. Their results involving the notion of common knowledge rather than common certainty, (see Fagin *et al*., 1995; Dekel & Gul, 1997) do not deal with strategic independence, and do not contain explicit assumptions concerning how players update 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 (1996) formalize the notion of "common certainty of admissibility" and thereby characterize an iterated deletion procedure which captures certain aspects of forward induction.

This paper is organized as follows. Section 2 discusses two examples which motivate our axiom systems and the solution concepts they characterize. Section 3 introduces the epistemic model. Weak and strong rationalizability, as well as the axioms which characterize them, are defined in Section 4, which contains the main results. Comments on explicability and results concerning backward induction are collected in Section 5. All proofs are in the Appendix.

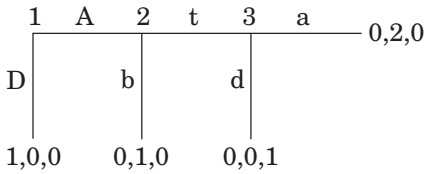
2. Motivating examples

As was anticipated in the introduction, the assumption that players choose their strategies independently suggests two related but distinct assumptions for extensive-game analysis.

The first pertains to players' beliefs in the course of the game. Quite naturally, in light of the assumption of causal independence between the strategies of different players, it seems interesting to explore the further assumption that players' conjectures exhibit some form of *epistemic (or stochastic) independence*. We would like to consider a restriction that in games of perfect information (like the examples below) has the following flavor: *a player will not revise her beliefs about an opponent until she observes an action taken*

by that particular opponent. Observations about *other* opponents' choices should be irrelevant.†

Subsection 4.1 introduces the formal and general notion of (epistemic) independence we adopt in this paper; the following example illustrates the basic intuition.



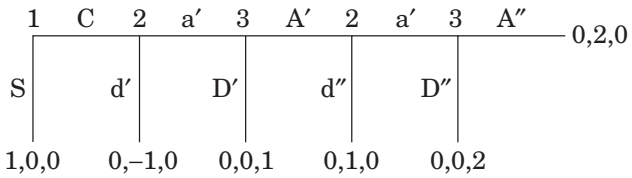
Notice that "D" is strictly dominant for player 1; also, "d" is conditionally strictly dominant for player 3. Hence, if player 2 is certain (at the beginning of the game) that 1 and 3 are sequentially rational, he should expect them to choose "D" and "d", respectively.

But what if player 2's node is actually reached? Player 2 cannot continue to believe that player 1 is rational: he has just received evidence to the contrary. If player 2 entertains the hypothesis that players 1 and 3 might be coordinating their strategies (e.g. if he believes that player 3 is really an "agent" of player 1) then he might be justified in expecting player 3 to choose "a" at the last node. Hence, in the absence of any independence assumption, we can justify player 2's choice of "t". Correlated EFR captures precisely this type of reasoning.

Once we require that players' beliefs be stochastically independent, predictions are narrowed down to a single strategy profile: player 2, upon observing 1's choice of "A", is *not* allowed to revise his initial beliefs about player 3. Thus, expecting "d" at the last node, he does well to choose "b".

Notice that, in order to reach this conclusion, we only need to assume that players are sequentially rational, have independent conjectures and are certain that this is the case *at the beginning of the game*: this is of course the type of restrictions characterizing *weak* rationalizability.

A slightly more complicated example illustrates the interaction between epistemic independence and the best rationalization principle.



† As Stalnaker (1996*b*) forcefully argues, causal independence does *not* entail epistemic independence.

It will be convenient to briefly describe the axioms proposed by Battigalli and Siniscalchi (1997) for comparison purposes. Let (0) = "Every player is (weakly) sequentially rational"; then, for $k > 0$, let (k) = "Every player strongly believes that (0)&(1)&...&($k-1$)". We may also consider a solution concept which is somehow halfway between extensive-form rationalizability and strong rationalizability; for this hybrid concept, (0I) = (0S) = "Every player is (weakly) sequentially rational and has epistemically independent beliefs" and, for $k > 0$, (kI) is defined analogously to (k).† Finally, weak rationalizability is characterized by the axioms (0W) = (0S), and for $k > 0$, (kW) = "Everybody is certain that (($k-1$)W) at the beginning of the game".

Let us now proceed with the analysis of the game. Observe first that player 1, if rational, will never play "C"; similarly, choosing "d′" is conditionally strictly dominated for player 2, and player 3 would choose "D′′" after "A′". Any remaining strategy of players 2 and 3 may be justified by some *independent* system of conditional beliefs over the strategy profiles of the respective opponents.

Hence, axioms (0) and (0W) = (0I) = (0S) identify the same set of strategy profiles. It is easy to see that, if we add axiom (1) to axiom (0), we do not obtain any further restrictions: as soon as player 2's first node is reached, the assumption that *every player* is rational is clearly falsified, so players 2 and 3 may update their beliefs *arbitrarily*. Hence, any rational strategy for these players survives the second (and therefore any successive) round of inductive reasoning.

Let us consider adding axiom (1W) to (0W) = (0I) = (0S). Player 2 assigns probability zero to player 3's irrational strategy "A′A′′" at the beginning of the game and epistemic independence implies that he continues to do so at his *first* node. However, the axioms do *not* pin down the conditional probability assigned by player 2 to "A′A′′" at his *second* node.

Specifically, suppose that player 2 initially was certain that player 3 would choose "D′". Upon reaching his second node, player 2 must conclude that his initial conjecture was wrong, and is therefore forced to form new beliefs. This clearly does not violate stochastic independence, as new information on player 3 has indeed been obtained. Moreover, since we are imposing rationality restrictions only on ex-ante beliefs, player 2's beliefs at his second node may be arbitrarily specified—which again allows one to justify any one of his rational strategies. It is easy to see that all of player 3's rational strategies may still be justified; hence, axiom (1W), too, imposes no further restrictions: weak and correlated rationalizability yield the same solution in this example.

† Recall that axioms (kS) ($k = 0, 1, \ldots$) are stated in the Introduction.

Next, consider adding axiom (1I) instead of (1W). Again, player 2 assigns probability zero to "A′A″" at his first node. But the key observation is that, even under this stronger axiom, player 2's updated beliefs at his second node are still *unrestricted*.

Again, suppose that his initial conjecture assigned marginal probability 1 to player 3's strategy "D′". This assumption is falsified at player 2's second node, so stochastic independence has no bite. Axiom (1I) requires player 2 to "strongly believe"—believe *whenever possible* —that *every player* is rational and has independent beliefs. While it is possible for player 2 to continue to believe that player 3 is rational (and has independent beliefs), he now is certain that player 1 is *not* rational. But this means that it is not possible for player 2 to believe that *everybody* is rational (and has independent beliefs), so in fact *axiom (1I) has no bite*. Thus, the three sets of axioms considered above induce the same solution in this game.

The example illustrates that we need to complement the epistemic independence assumption with an additional restriction on *belief revision*. The intuitive notion we would like to capture is the following: *each player separately assesses the degree of strategic sophistication of every one of her opponents*. By contrast, correlated rationalizability reflects the assumption that players assess the *joint* degree of strategic sophistication of their opponents, viewed as a *group*; more specifically, players attribute to each one of their opponents (at least) the degree of strategic sophistication of the "least sophisticated" among them.

In the current example, axiom (1S.2) restricts player 2's beliefs at his second node: he must necessarily expect player 3 to follow "A′" with "D″", because "A′A″" is her only strategy which reaches player 2's second node and is consistent with axiom (0S.3). This is true even if axiom (0S.1) clearly cannot hold if the node under consideration is reached.

It is now easy to see that axiom (2S.3) implies that player 3 will anticipate this, and hence choose "D′" at her first node. Strong rationalizability thus yields more stringent restrictions than the other solution concepts considered here (although, in this particular example, it selects the same outcome).

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. We use the following notation:

$i \in N = \{1, \ldots, n\}$, players;
$h \in \mathcal{H}_i$, information sets for player i;

$s_i \in S_i$, pure strategies for player i;
$S = \prod_{i=1}^{n} S_i$, $S_{-i} = \prod_{j \neq i} S_j$;
$U_i : S \to \mathbf{R}$, strategic-form payoff function for player i;
$s \in S(h)$, strategy-profiles reaching $h \in \bigcup_{i \in I} \mathcal{H}_i$;
$\mathcal{H}_i(s_i) = \{h \in \mathcal{H}_i : \exists s_{-i} \in S_{-i}$ 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.

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 & Rubinstein, 1994, Ch. 6). For example, in multi-stage 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.†

3.2. EPISTEMIC MODEL

3.2.1. Conditional probability systems

Consider a collection of Polish (complete, separable, metrizable) spaces $\{Y_1, \ldots, Y_n\}$. We interpret $y_i \in Y_i$ as an *unobservable* (and payoff irrelevant) parameter representing the conditional beliefs of player i. The Cartesian product $\prod_{i=1}^{n} S_i \times Y_i$ is also Polish (we endow each S_i with the discrete topology and $\prod_{i=1}^{n} S_i \times Y_i$ with the product topology). Let $\mathcal{S} = 2^S \setminus \{\emptyset\}$ denote the collection of all the non-empty subsets of S. Fix a non-empty collection $\mathcal{B} \subseteq \mathcal{S}$ of "relevant hypotheses" about $s \in S$. Then we obtain a corresponding collection,

$$\mathcal{C}(\mathcal{B}) = \left\{ C \subseteq \prod_{i=1}^{n} S_i \times Y_i : \exists B \in \mathcal{B}, C = \{(s_i, t_i)_{i \in N} : (s_1, \ldots, s_n) \in B\} \right\}$$

of "relevant hypotheses" about $(s_i, y_i)_{i \in N} \in \prod_{i=1}^{n} S_i \times Y_i$. Let $\mathcal{A}$ be the Borel σ-algebra on $\prod_{i=1}^{n} S_i \times Y_i$. Clearly $\mathcal{C}(\mathcal{B}) \subset \mathcal{A}$. A *conditional probability system* (or CPS) on $(\prod_{i=1}^{n} S_i \times Y_i, \mathcal{A}, \mathcal{B})$ is a map,

$$\mu(\cdot|\cdot) : \mathcal{A} \times \mathcal{C}(\mathcal{B}) \to [0, 1]$$

† Battigalli and Siniscalchi (1998) analyze multistage games with observed actions and incomplete information.

satisfying the following axioms:

AXIOM 1: *For all* $B \in \mathcal{C}(\mathcal{B})$, $\mu(\cdot|B)$ *is a probability measure on* $(\prod_{i=1}^{n} S_i \times Y_i, \mathcal{A})$.

AXIOM 2: *For all* $B \in \mathcal{C}(\mathcal{B})$, $\mu(B|B) = 1$.

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

The set of probability measures on a measure space $(Z, \mathcal{A})$ is denoted by $\Delta(Z)$; the set of conditional probability systems on $(\prod_{i=1}^{n} S_i \times Y_i, \mathcal{A}, \mathcal{B})$ can be regarded as a subset of $[\Delta(\prod_{i=1}^{n} S_i \times Y_i)]^{\mathcal{B}}$ (the set of mappings from $\mathcal{B}$ to $\Delta(\prod_{i=1}^{n} S_i \times Y_i)$) and it is denoted by $\Delta^{\mathcal{B}}(\prod_{i=1}^{n} S_i \times Y_i)$. ($\Delta^{\mathcal{B}}(S)$ is similarly defined.) Accordingly, we often write $\mu = (\mu(\cdot|B \times Y))_{B \in \mathcal{B}} \in \Delta^{\mathcal{B}}(\prod_{i=1}^{n} S_i \times Y_i)$. The topology on $\prod_{i=1}^{n} S_i \times Y_i$ and $\mathcal{A}$, the corresponding Borel σ-algebra, are usually understood and need not be explicit in our notation. Thus we simply refer to "conditional probability system (or CPS) on $(\prod_{i=1}^{n} S_i \times Y_i, \mathcal{B})$". It is also understood that $\Delta(\prod_{i=1}^{n} S_i \times Y_i)$ is endowed with the topology of weak convergence of measures and $[\Delta(\prod_{i=1}^{n} S_i \times Y_i)]^{\mathcal{B}}$ is endowed with the product topology. Thus $\Delta(\prod_{i=1}^{n} S_i \times Y_i)$ and $[\Delta(\prod_{i=1}^{n} S_i \times Y_i)]^{\mathcal{B}}$ (by countability of $\mathcal{B}$) are Polish spaces. Since $\Delta^{\mathcal{B}}(\prod_{i=1}^{n} S_i \times Y_i)$ is a closed subset of $[\Delta(\prod_{i=1}^{n} S_i \times Y_i)]^{\mathcal{B}}$, also $\Delta^{\mathcal{B}}(\prod_{i=1}^{n} S_i \times Y_i)$ is a Polish space (endowed with the relative topology inherited from $[\Delta(\prod_{i=1}^{n} S_i \times Y_i)]^{\mathcal{B}}$).

3.2.2. *Universal type space*

For any measurable product space $X \times Y$ and any probability measure $\mu \in \Delta(X \times Y)$ let $mrg_X \mu \in \Delta(X)$ denote the marginal of μ on X. A *universal type space on* $(S_i, \mathcal{B}_i)_{i \in N}$ is given by a tuple $(T_i, g_i)_{i \in N}$ whereby, for every player $i \in N$, T_i is a Polish space, the function

$$g_i = (g_{i,B})_{B \in \mathcal{B}_i} : T_i \to \Delta^{\mathcal{B}_i}\left(\prod_{i=1}^{n} S_i \times T_i\right)$$

satisfies

$$\forall B \in \mathcal{B}_i, \forall t_i \in T_i, g_{i,B}(t_i)\left(\left\{(s'_j, t'_j)_{j \in N} : t'_i = t_i\right\}\right) = 1$$

and the corresponding function $\overline{g}_i$ defined by

$$t_i \longmapsto \left(mrg_{S_1 \times T_1 \times \cdots \times S_i \times \cdots \times S_n \times T_n} g_{i,B}(t_i)\right)_{B \in \mathcal{B}_i}$$

is a *homeomorphism* between T_i and $\Delta^{\mathcal{B}_i}(S_1 \times T_1 \times \cdots \times S_i \times \cdots \times S_n \times T_n)$. We shall denote the Borel σ-algebra on $\prod_{i \in N}(S_i \times T_i)$ by $\mathcal{E}$.

Universal type spaces of this sort are explicitly constructed and analyzed by Battigalli and Siniscalchi (1998) to which we refer for details.† An element $t_i \in T_i$ represents a possible *epistemic type* for player i. The equation above is an introspection property essentially saying that a player knows his own type.

REMARK 1: For all $i \in N$, $t_i \in T_i$, $p \in [0, 1]$, $B \in \mathcal{B}_i$ and measurable subsets $E \subseteq \prod_{i=1}^{n} S_i \times Y_i$, if

$$g_{i,B}(t_i)(E) \geq p,$$

then, for all $C \in \mathcal{B}_i$,

$$g_{i,C}(t_i)\left(\left\{(s'_j, t'_j)_{j \in N} : g_{i,B}(t'_i)(E) \geq p\right\}\right) = 1.$$

Note that, in general, distinct players may have different collections of relevant hypotheses. In fact, the most natural collection for player i is the one representing his information sets:

$$\mathcal{B}_i^{\mathcal{H}} = \{B \subseteq S : \exists h \in \mathcal{H}_i, B = S(h)\}.$$

However, in this paper we will assume that players regard any non-empty subset of S (that is, any event regarding the players' dispositions to act, but not their epistemic states) as a relevant conditioning hypothesis: that is, we take $\mathcal{B}_i = \mathcal{S}$ for all $i \in N$. Of course, this implies that type spaces are (almost entirely) *symmetric*.‡

An epistemic type corresponds to an infinite *hierarchy of conditional beliefs*. In fact, for every $i \in N$ and $t_i \in T_i$, we can derive the marginal CPS $\varphi_i^1(t_i) = (mrg_{S^{\mathcal{B}}} g_{i,B}(t_i))_{B \in \mathcal{B}_i}$, that is, the *first order* conditional beliefs of type t_i. The maps $\varphi_j^1 : T_j \to \Delta^{\mathcal{B}_j}(S)$ $(j \in N)$ can be used to derive the *second order* CPS $\varphi_i^2(t_i)$ implicit in t_i: for each measurable set $A^1 \subseteq \prod_{j \in N} S_j \times \Delta^{\mathcal{B}_j}(S)$ and relevant hypothesis $B \in \mathcal{B}_i$,

$$\varphi_{i,B}^2(t_i)(A^1) = g_{i,B}(t_i)\left(\{(s'_j, t'_j) : [s'_j, \varphi_j^1(t'_j)]_{j \in N} \in A^1\}\right);$$

then let $\varphi_i^2(t_i) = (\varphi_{i,B}^2(t_i))_{B \in \mathcal{B}_i}$. Similarly, using the maps φ_j^1 and $\varphi_j^2 : T \to \Delta^{\mathcal{B}_j}(\prod_{k \in N}[S_k \times \Delta^{\mathcal{B}_k}(S)])$ $(j \in N)$ we can derive the *third*

† Battigalli and Siniscalchi (1998) consider symmetric type spaces for two-person dynamic games and do not explicitly represent a player's conditional beliefs about his own conditional beliefs, but the appropriate modifications to their analysis to fit the present framework are straightforward.

‡ Even in this "almost symmetric" case, the sets of types and belief functions of distinct players are formally different, because they satisfy different introspection properties.

order CPS $\varphi^3(t_i)$ implicit in type t_i: for each measurable set

$$A^2 \subseteq \prod_{j \in N} S_j \times \Delta^{\mathcal{B}_j}(S) \times \Delta^{\mathcal{B}_j}\left(\prod_{k \in N} S_k \times \Delta^{\mathcal{B}_k}(S)\right),$$

and relevant hypothesis $B \in \mathcal{B}_i$,

$$\varphi^3_{i,B}(t_i)(A^2) = g_{i,B}(t_i)\left(\{(s'_j, t'_j) : [s'_j, \varphi^1_j(t'_j), \varphi^2(t'_j)]_{j \in N} \in A^2\right);$$

then let $\varphi^3_i(t_i) = (\varphi^3_{i,B}(t_i))_{B \in \mathcal{B}_i}$. In general, we can derive an infinite hierarchy of CPSs $\varphi_i(t_i) = \varphi^1_i(t_i), \varphi^2_i(t_i), \ldots$ with an inductive procedure. Furthermore, for each type t_i the corresponding infinite hierarchy of CPSs $\varphi_i(t_i)$ satisfies a natural coherency condition; let X^k be the space over which $(k+1)$-order beliefs are defined (thus, $X^0 = S, X^1 = \prod_{j \in N} S_j \times \Delta^{\mathcal{B}_j}(S), \ldots$): then, for all $k = 1, 2, \ldots, B \in \mathcal{B}_i$,

$$mrg_{X^{k-1}} \varphi^{k+1}_{i,B}(t_i) = \varphi^k_{i,B}(t_i).$$

Since this holds for every type of every player, every type of every player is certain conditional on every relevant hypothesis that everybody's type satisfies the coherency condition, every type of every player is certain, conditional on every relevant hypothesis, of the latter fact and so on. The same holds for the introspection property relative to beliefs of any order.

We call the type space $(T_i, g_i)_{i \in N}$ *universal* because it can be shown that, since each associated function $\overline{g}_i$ is a homeomorphism, each T_i "contains" every infinite hierarchy of CPSs satisfying conditional common certainty of coherency and introspection. This means that focusing on such a type space we are not implicitly introducing extraneous assumptions about players' conditional beliefs of any order. As argued in Battigalli and Siniscalchi (1997), this is crucial if we are to provide a transparent epistemic characterization of extensive form solution concepts involving some form of forward induction.

4. Epistemic independence and rationalizability

In this section we define and characterize a weak and a strong notion of extensive form rationalizability with independent beliefs. The weak notion of rationalizability is a refinement of a solution concept proposed by Ben Porath (1997) to characterize common certainty of rationality in extensive form games. The strong notion of rationalizability has been put forward by Battigalli (1996) to amend a flaw in Pearce's notion of extensive form rationalizability (see Pearce, 1994) and is related to an iterative deletion procedure proposed by Reny (1992) to define explicable equilibrium, a

refinement of the Nash equilibrium concept. The relationship between each of these solution concepts and backward induction is noted below. The notion of independence used here is related to, but weaker than Kreps and Wilson's (1982) consistency of assessments and has been given a decision theoretic axiomatization by Battigalli and Veronesi (1996). The phrase "epistemic independence" is borrowed from Stalnaker (1996*b*) who uses it to indicate a very similar property.

Recall that we assume here that, for each player i, the collection of relevant hypotheses coincides with the whole set of non-empty subsets of S, or—in our notation—$\mathcal{B}_i = \mathcal{S}$ for all $i \in N$. This simplifies our notation and allows a transparent definition of the independence property for general games.†

To make our formulation simpler, the following notation will be convenient. A first order CPS for an arbitrary player i is typically denoted by δ_i, that is, $\delta_i \in \Delta^{\mathcal{S}}(S)$. For every epistemic type $t_i \in T_i$ and hypothesis $B \in \mathcal{S}$, $\delta_{i,B}(t_i)$ denotes the marginal of $g_{i,B}(t_i)$ on S. Clearly, $\delta_i(t) \equiv (\delta_{i,B}(t))_{B \in \mathcal{S}} = \varphi_i^1(t_i)$ is a (complete) first order CPS on S. For any non-empty group of players $J \subset N$,‡ $\mathcal{E}_J$ denotes the Borel σ-algebra on $\prod_{j \in J} S_j \times T_j$ with typical elements $A_J, B_J, \ldots$. The collection of all non-empty subsets of $S_J = \prod_{j \in J} S_j$ is denoted $\mathcal{S}_J = 2^{S_J} \setminus \{\emptyset\}$ and $\mathcal{C}_J(\mathcal{S}_J)$ denotes the collection of "cylinders" in $\mathcal{E}_J$ with base $\mathcal{S}_J$. For any product space $\prod_{i=1}^n X_i$, for all $\emptyset \neq J \subset N$, $A_J \subseteq \prod_{j \in J} X_j$, $B_{N \setminus J} \subseteq \prod_{k \in N \setminus J} X_k$, we abuse notation slightly and write

$$A_J \times B_{J \setminus K} = \left\{ (x_i)_{i \in N} : (x_j)_{j \in J} \in A_J, (x_k)_{k \in N \setminus J} \in B_{N \setminus J} \right\}.$$

Again with an abuse of notation, we do not distinguish between singletons and their unique element whenever the meaning is clear from the context. For any CPS $\mu_i \in \Delta^{\mathcal{S}}(\prod_{j=1}^n S_j \times T_j)$ and $\emptyset \neq J \subset N$, the *marginal* of μ_i on $\prod_{j \in J} S_j \times T_j$ is the function $\mu_{iJ}(\cdot|\cdot) : \mathcal{E}_J \times \mathcal{C}_J(\mathcal{S}_J) \to [0, 1]$ defined by the following equalities: for all $A_J \in \mathcal{E}_J$, $B_J \in \mathcal{C}_J(\mathcal{S}_J)$,

$$\mu_{iJ}(A_J|B_J) = \mu_i\left(A_J \times \left(\prod_{k \in N \setminus J} S_k \times T_k\right) \middle| B_J \times \left(\prod_{k \in N \setminus J} S_k \times T_k\right)\right).$$

It is easily checked that μ_{iJ} is a CPS on $(\prod_{j \in J} S_j \times T_j, \mathcal{S}_J)$. An analogous (and simpler) definition holds for the marginal δ_{iJ} of a first order CPS $\delta_i \in \Delta^{\mathcal{S}}(S)$.

† For games with observable deviators (which include games with observable actions), we can provide a more natural and parsimonious formulation whereby each $\mathcal{B}_i = \mathcal{B}_i^H$.

‡ We use the symbol $\subset$ to denote *strict* inclusion.

4.1. EPISTEMIC INDEPENDENCE

A CPS $\mu_i \in \Delta^{\mathcal{S}}(\prod_{j=1}^{n} S_j \times T_j)$ (and the corresponding type $t_i = (g_i)^{-1}(\mu_i)$) satisfies the *epistemic independence* property if the marginal conditional beliefs about any group of players J are unaffected by information exclusively concerning the complementary group $N \backslash J$, that is, for all $\emptyset \neq J \subset N$, $A_J \in \mathcal{E}_J$, $B_J \in \mathcal{C}_J(\mathcal{S}_J)$, $C_{N\backslash J} \in \mathcal{C}_{N\backslash J}(\mathcal{S}_{N\backslash J})$,

$$\mu_i(A_J \times C_{N\backslash J} | B_J \times C_{N\backslash J}) = \mu_{iJ}(A_J | B_J).$$

An analogous definition holds for first order conditional systems $\delta \in \Delta^{\mathcal{S}}(S)$. Let $I\Delta^{\mathcal{S}}(\prod_{j=1}^{n} S_j \times T_j)$ and $I\Delta^{\mathcal{S}}(S)$ be the sets of CPSs (on $(\prod_{j=1}^{n} S_j \times T_j, \mathcal{S})$ and $(S, \mathcal{S})$ respectively) satisfying the epistemic independence condition. For brevity, we simply call such conditional systems *independent*.

REMARK 2: *For all $i \in N$, $t_i \in T_i$, if $g_i(t_i)$ is independent, then $\delta_i(t_i)$ (the first order CPS of t_i) is also independent.* We show in the Appendix (see Lemma 2) that a sort of "converse" is also true: if a given first order CPS $\overline{\delta}_i$ is independent, then there exists an independent type t_i such that $\overline{\delta}_i = \delta_i(t_i)$.

4.2. WEAK AND STRONG RATIONALIZABILITY WITH INDEPENDENT BELIEFS

The basic building block of the following solution concepts is the notion of *weak sequential rationality*. This is a best response property which applies to plans of action† as well as strategies (see e.g. Reny, 1992). We adopt the specific formalization proposed in Battigalli and Siniscalchi (1998) (see their Definition 5.1):

DEFINITION 1: *Fix a first order CPS $\delta_i \in \Delta^{\mathcal{S}}(S)$. A strategy $s_i \in S_i$ is a weakly sequential best reply to δ_i if, 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 Battigalli & Siniscalchi (1998) for details on the features of this definition. Here we simply point out that part 1 essentially means that a rational player is certain

† 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.

of her strategy (hence of her future contingent choices) as long as she knows that she has not deviated from it. If δ_i is independent, as assumed below, then the relevant conditional beliefs about i's opponents are given by $\delta_{i,-i}$ (the marginal of δ_i on S_{-i}) and we require that s_i be a best response to $\delta_{i,-i}(\cdot|S_{-i}(h))$ 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 $r_i : \Delta^{\mathcal{S}}(S) \to S_i \cup \{\emptyset\}$ assigning to each $\delta_i \in \Delta^{\mathcal{S}}(S)$ the unique weakly sequential best reply s_i or the symbol $\emptyset$, as the case may be. Also, for every epistemic type $t_i \in T_i$, we let $\rho_i(t) = r_i[\delta_i(t_i)]$ denote the "best reply" to the first order beliefs of type t_i.

We are now ready to define the weak and strong rationalizability procedure.

DEFINITION 2: *(cf. Ben Porath, 1997) For every $i \in N$, let $W_i^0 = S_i$. For every $k > 0$, $i \in N$ and $s_i \in S_i$, say that $s_i \in S_i^k$ if (and only if) there exists an independent CPS $\delta_i \in I\Delta^{\mathcal{S}}(S)$ such that:*

(1) $s_i = r_i(\delta_i)$,
(2) $\delta_i(\prod_{j \in N} W_j^{k-1}|S) = 1$.

Clearly, $W_i^{k+1} \subseteq W_i^k$ for all i and k. A strategy s_i for player i is *weakly rationalizable* if $s_i \in \bigcap_{k>0} W_i^k$.

DEFINITION 3: *(cf. Battigalli, 1996) For every $i \in N$, let $S_i^0 = S_i$. For every $k > 0$, $i \in N$ and $s_i \in S_i$, say that $s_i \in S_i^k$ if there exists an independent CPS $\delta_i \in I\Delta^{\mathcal{S}}(S)$ such that:*

(1) $s_i = r_i(\delta_i)$,
(2) For every $j \neq i$ and $s_j, s_j' \in S_j$: if $s_j \in S_j^m \setminus S_j^{m'}$, $s_j' \in S_j^{m'}$ and $k - 1 \geq m' > m$, then $\delta_{ij}(s_j'|s_j, s_j') = 1$.

Note that, by condition (2) of Definition 3, $S_i^{k+1} \subseteq S_i^k$ for all i and k. A strategy s_i for player i is *strongly rationalizable* if $s_i \in \bigcap_{k>0} S_i^k$.

Battigalli (1996) shows that the set of strongly rationalizable strategies is non-empty. The following shows that our terminology is consistent and implies that the set of weakly rationalizable strategies is also non-empty.

PROPOSITION 1: *Strong rationalizability implies weak rationalizability, that is, for all $i \in N$, $k = 1, 2, \ldots$, $S_i^k \subseteq W_i^k$.*

4.3. EPISTEMIC CHARACTERIZATION

In our epistemic model, the set of *states of the world* is the product space $\prod_{i=1}^{n}(S_i \times T_i)$. An *event* is a measurable subset $E \subseteq \prod_{i=1}^{n}(S_i \times T_i)$.

However, we will often be interested in events which concern a given player j exclusively. In keeping with our notation, $\mathcal{E}_j$ denotes the Borel σ-algebra on $S_j \times T_j$. For any $E \in \mathcal{E}_j$, let

$$[E]_j = \{(s_j, s_{-j}, t_j, t_{-j}) \ : \ (s_j, t_j) \in E\}$$

That is, $[E]_j$ is the cylinder in $\prod_{i\in N}(S_i \times T_i)$ with base $E \subseteq S_j \times T_j$. We shall say that E *corresponds to* the event $[E]_j$, which concerns player j only.

Thus the subsets

$$R_j = \left\{(s_j, t_j) : s_j = \rho_j(t_j)\right\}$$

and

$$I_j = \left\{(s_j, t_j) : g_j(t_j) \in I\Delta^{\mathcal{S}}\left(\prod_{i=1}^{n} S_i \times T_i\right)\right\}$$

respectively correspond to the events "player j is rational" and "player j has independent beliefs". $R = \prod_{i=1}^{n} R_i$ and $I = \prod_{i=1}^{n} I_i$ are the events "everyone is rational" and "everyone has independent beliefs".†

4.3.1. Weak rationalizability

We say that an epistemic type t_j is *certain* of event E if the *prior* belief of t_i assigns probability 1 to E, that is, $g_{j,S}(t_j)(E) = 1$. Let

$$\beta_j(E) = \left\{(s_j, t_j) : g_{j,S}(t_j)(E) = 1\right\}$$

be the subset corresponding to the event "player j is certain of E"; then

$$\beta(E) = \prod_{i=1}^{n} \beta_i(E)$$

is the event "everybody is certain of E". β has all the standard properties of a common belief operator‡, but of course it does

† Observe that these events could also be written as $R = \bigcap_{i\in N}[R_i]_i$ and $I = \bigcap_{i\in N}[I_i]_i$ respectively, but the notation used in the text is somewhat more suggestive.

‡ That is, consistency $[\beta(\emptyset) = \emptyset]$, logical omniscience $[\beta(\prod_{i=1}^{n} S_i \times T_i) = \prod_{i=1}^{n} S_i \times T_i]$, conjunctiveness $[\beta(E) \cap \beta(F) \subset \beta(E \cap F)]$ and monotonicity $[E \subseteq F \Rightarrow \beta(E) \subseteq \beta(F)]$.

not satisfy the truth axiom: $\beta(E) \subseteq E$ does *not* necessarily hold. Iterations of β are defined in the usual way and denote mutual certainty of degree k. Let $\beta^0(E) = E$ by convention. Then for all $k = 1, 2 \ldots$,

$$\beta^k(E) = \beta(\beta^{k-1}(E)).$$

Thus, the event "it is the case that E and there is common certainty of E" is

$$E \cap \left(\bigcap_{k \geq 1} \beta^k(E) \right) = \bigcap_{k \geq 0} \beta^k(E).$$

PROPOSITION 2: *For all strategy profiles $s = (s_i)_{i \in N} \in S$, the following statements hold:*

(1) For all $k = 0, 1, \ldots$, $s \in \prod_{i=1}^n W_i^{k+1}$ if and only if there exists a profile of epistemic types $(t_i)_{i \in N}$ such that $(s_i, t_i)_{i \in N} \in \bigcap_{m=0}^k \beta^m(I \cap R)$.

(2) $s \in \prod_{i=1}^n (\bigcap_{k \geq 1} W_i^k)$ if and only if there is a profile of epistemic types $(t_i)_{i \in N}$ such that $(s_i, t_i)_{i \in N} \in \bigcap_{k \geq 0} \beta^k(I \cap R)$.

4.3.2. Strong rationalizability

We now move on to strong rationalizability. The following definition introduces the first key ingredient in our axiomatization. We formalize the idea that a player i may formulate a conjecture about a particular opponent j, and for every hypothesis concerning j only and consistent with such conjecture she may be unwilling to revise it.

Recall that $\mathcal{S}_j = 2^{S_j} \setminus \{\emptyset\}$, which we view as the counterpart to $\mathcal{E}_j$ in S_j.

DEFINITION 4: *For any pair of players $i, j \in N$, type $t_i \in T_i$ and measurable subset $E_j \subseteq S_j \times T_j$, we say that type t_i strongly believes E_j if for all $B_j \in \mathcal{S}_j$,*

$$E_j \cap (B_j \times T_j) \neq \emptyset \Rightarrow g_{i,B_j \times S_{-j}}(t_i)([E_j]_j) = 1.$$

Let $\beta_{ij}^*(E) \in \mathcal{E}_i$ correspond to the event that player i strongly believes $E_j \in \mathcal{E}_j$; formally,

$$\beta_{ij}^*(E_j) := \left\{ (s_i, t_i) : \forall B_j \in \mathcal{S}_j, E_j \cap (B_j \times T_j) \neq \emptyset \Rightarrow g_{i,B_j \times S_{-j}}(t_i)([E_j]_j) = 1 \right\}$$

The second ingredient in our axiomatization is the *independent best rationalization principle*. The idea (which will be made explicit in Remark 3 below) is that, at each point in the game, players bestow the highest possible degree of strategic sophistication upon

each one of their opponents independently. That is, for all i and j ($j \neq i$), player i strongly believes that j is rational, *and* for all i, j and k ($j \neq i, k \neq j$), i strongly believes that j strongly believes that k is rational, etc.

We formalize the best rationalization principle as follows. Given any collection $(E_j)_{j\in N} \in \mathcal{E}_1 \times \mathcal{E}_2 \times \ldots \times \mathcal{E}_n$, for every $i \in N$, let

$$\gamma_i^0[(E_j)_{j\in N}] = E_i;$$

then, for $l > 0$, let

$$\gamma_i^l[(E_j)_{j\in N}] = \gamma_i^{l-1}[(E_j)_{j\in N}] \cap \bigcap_{j\neq i} \beta_{ij}^* \left(\gamma_j^{l-1}[(E_k)_{k\in N}]\right)$$

Clearly, $\gamma_i^l[(E_j)_{j\in N}] \in \mathcal{E}_i$ for every $i \in N$ (so the above definition is indeed meaningful). Also, the sets $\gamma_i^l(.)$ form a monotonically decreasing sequence.

The following remark further clarifies the nature of our assumptions. It also justifies the informal statement of the axioms given in the Introduction.

REMARK 3: For every $l \geq 0$ and $i \in N$,

$$\gamma_i^l((E_j)_{j\in N}) = E_i \cap \bigcap_{m=0}^{l-1} \bigcap_{j\neq i} \beta_{ij}^*(\gamma_j^m((E_k)_{k\in N}).$$

We can finally state the characterization result:

PROPOSITION 3: *For all strategy profiles $s = (s_i)_{i\in N} \in S$ the following statements hold:*

(1) For all $k = 0, 1, \ldots, s \in \prod_{i=1}^n S_i^{k+1}$ if and only if there exists a profile of epistemic types $(t_i)_{i\in N}$ such that $(s_i, t_i)_{i\in N} \in \prod_{i=1}^n \gamma_i^k[(R_j \cap I_j)_{j\in N}]$.

(2) $s \in \prod_{i=1}^n (\bigcap_{k\geq 1} S_i^k)$ if and only if there is a profile of epistemic types $(t_i)_{i\in N}$ such that $(s_i, t_i)_{i\in N} \in \prod_{i=1}^n \bigcap_{k\geq 0} \gamma_i^k[(R_j \cap I_j)_{j\in N}]$.

5. Relationship with other solution concepts

5.1. EXPLICABILITY

As we have mentioned above, strong rationalizability and explicability (Reny, 1992) are similarly motivated and formally comparable. This is apparent if one compares our Definition 3 with the iterative procedure defined by Reny (1992, p. 639).

Analogously to strong rationalizability, Reny's procedure identifies a partition of each player's strategy space whose elements are ordered according to their degree of strategic sophistication (or "explicability"). A set of "beliefs" (see below) is associated with each element of the partition; beliefs associated with the set of "k-th order explicable" strategies satisfy a restriction analogous to condition (2) in Definition 3. Finally, strategies are $k+1$-th order explicable if they are weakly sequential best replies relative to beliefs consistent with k-th order explicability.

However, explicability and strong rationalizability are not equivalent.

First, Reny models beliefs by means of *consistent assessments*, as in Kreps and Wilson (1982). As is well known, this entails a strong notion of independence—stronger than our epistemic independence condition. However, for games with *observable deviators* this difference disappears (see e.g. Battigalli, 1996 and references therein).

The second difference is more subtle, but substantive and not easily circumvented. Given a family of behavioral strategy profiles, Reny deems a strategy a best reply relative to that family if it can be rationalized by *any element in the convex hull of that set*.

Our setup cannot accommodate such an assumption in a natural way. Indeed, we are suspicious about its legitimacy.† Note that Selten (1975) explicitly avoids taking pointwise convex combinations of behavioral strategies by defining a notion of *behavioral strategy mixtures*, whereby different behavioral strategies may be selected *before the game begins* according to a random mechanism and conditional probabilities of actions are then derived *via* Bayes rule. For example, move probabilities at information sets which can be reached by only one of the behavioral strategies, among which the mechanism choises are determined by that strategy only. It should be clear that defining beliefs as CPS's on the set of *pure* strategy profiles, as we do, allows for behavioral strategy mixtures.

5.2. BACKWARD INDUCTION

We now turn to the relationship between each of the solution concepts characterized in the previous section and backward

† Suppose that in a two-person, two-stage game player 1 has only two rational (pure) plans of action: L followed by L or R followed by R. Clearly, if player 2 strongly believes that player 1 is rational, he should expect in the second stage the same action observed in the first. But taking pointwise convex combinations of rational behavioral strategies, one typically obtains non-degenerate expectations in the second stage.

induction. In particularly simple games, weak rationalizability is sufficient to yield the backward induction solution:

PROPOSITION 4: *Suppose that the given game has perfect information, no player moves more than once in any path and there are no ties between payoffs at terminal nodes. Then, for all $i \in N$, there is a unique weakly rationalizable strategy s_i and $(s_i)_{i \in N}$ is the unique subgame-perfect equilibrium.*

Hence, within this class of games, we can provide relatively weak epistemic conditions for subgame perfection:

COROLLARY 1: *Suppose the given game satisfies the assumptions of Proposition 4. If $(s_i, t_i)_{i \in N} \in \bigcap_{k \geq 0} \beta^k(I \cap R)$ then $(s_i)_{i \in N}$ is the unique subgame-perfect equilibrium.*

Next, we consider strong rationalizability in perfect information games with arbitrary structure; we maintain the assumption that payoffs are generic. The following result parallels a property of explicable equilibria (Reny, 1992) and is proved in an entirely similar fashion.†

However, observe that, as discussed above, explicability and strong rationalizability entail different assumptions on beliefs (even in perfect information games), so one cannot simply invoke Reny's results.

PROPOSITION 5: *Suppose that the given game has perfect information and there are no ties between payoffs at terminal nodes. Then all strategy profiles $(s_i)_{i \in N} \in \prod_{i \in N} \bigcap_{k \geq 0} S_i^k$ induce the same path, which coincides with the unique subgame-perfect equilibrium path.*

Sufficient epistemic conditions for backward induction may then be easily given:

COROLLARY 2: *Suppose that the given game satisfies the assumptions of Proposition 4. If $(s_i, t_i)_{i \in N} \in \prod_{i \in N} (\bigcap_{k \geq 0} \gamma_i^k((R_j \cap I_j)_{j \in N}))$, then the strategy profile $(s_i)_{i \in N}$ induces the same path as the unique subgame-perfect equilibrium.*

We caution the reader that the result only guarantees that the backward induction *outcome* will obtain. The supporting beliefs may well differ from those prescribed by subgame perfection: for an example, see Battigalli and Siniscalchi (1997), Section 4.2.

† The (rather lengthy) argument is similar to Reny's proof of Proposition 3 in Reny (1992). While there are technical differences, the key ideas are essentially the same. The proof is available from the authors upon request.

References

Asheim, G.B. & M. Dufwenberg (1996). *Admissibility and common knowledge*. Department of Economics, University of Oslo, mimeo.

Aumann, R.J. (1995). Backward induction and common knowledge of rationality. *Games and Economic Behavior*, **8**, 6–19.

Aumann, R.J. (1996). Reply to binmore. *Games and Economic Behavior*, **17**, 138–146.

Aumann, R.J. (1998). On the centipede game: note. *Games and Economic Behavior*, **23**, 97–105.

Battigalli, P. (1996) Strategic independence and perfect bayesian equilibria. *Journal of Economic Theory*, **70**, 201–234.

Battigalli, P. (1996). Strategic rationality orderings and the best rationalization principle. *Games and Economic Behavior*, **13**, 178–200.

Battigalli, P. (1997). On rationalizability in extensive games. *Journal of Economic Theory*, **74**, 40–60.

Battigalli, P. & Siniscalchi, M. (1997). An epistemic characterization of extensive form rationalizability. *Social Sciences Working Paper*, 1009, Caltech.

Battigalli, P. & Siniscalchi, M. (1998). *Hierarchies of conditional beliefs and interactive epistemology in dynamic games*. Princeton University and European University Institute, Florence, mimeo.

Battigalli, P. & Veronesi, P. (1996). A note on stochastic independence without savage-null events. *Journal of Economic Theory*, **70**, 235–248.

Ben Porath, E. (1997). Rationality, nash equilibrium and backwards induction in perfect information games. *Review of Economic Studies*, **64**, 23–46.

Bernheim, D. (1984). Rationalizable strategic behavior. *Econometrica*, **52**, 1002–1028.

Brandenburger, A. & Dekel, E. (1993). Hierarchies of beliefs and common knowledge. *Journal of Economic Theory*, **59**, 189–198.

Dekel, E. & Gul, F. (1997). Rationality and knowledge in game theory. In D. Kreps & K. Wallis, Eds. *Advances in Economics and Econometrics*. Cambridge: Cambridge University Press.

Fagin, R., Halpern, J., Moses, J. & Vardi, M. (1995). *Reasoning about knowledge*. Cambridge MA: M.I.T. Press.

Kohlberg, E. & Mertens, J.F. (1986). On the strategic stability of equilibria. *Econometrica*, **54**, 1003–1037.

Kreps, D. & Wilson, R. (1982). Sequential equilibria. *Econometrica*, **50**, 863–894.

Myerson, R. (1986). Multi-stage games with communication. *Econometrica*, **54**, 323–358.

Osborne, M. & Rubinstein, A. (1994). *A Course in game theory*. Cambridge MA: MIT Press.

Pearce, D. (1984). Rationalizable strategic behavior and the problem of perfection. *Econometrica*, **52**, 1029–1050.

Reny, P. (1985). Rationality, common knowledge and the theory of games. Department of Economics, Princeton University, mimeo.

Reny, P. (1992). Backward induction, normal form perfection and explicable equilibria. *Econometrica*, **60**, 626–649.

Rênyi, A. (1956). On conditional probability spaces generated by a conditionally ordered set of measures. *Theory of Probability and Its Applications*, **1**, 61–71.

Samet, D. (1996). Hypothetical knowledge and games with perfect information. *Games and Economic Behavior*, **17**, 230–251.

Selten, R. (1975). Re-examination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory*, **4**, 25–55.

Stalnaker, R. (1996*a*). Knowledge, belief and counterfactual reasoning in games. *Economics and Philosophy*, **12**, 133–163.

Stalnaker, R. (1996*b*). Belief revision in games: forward and backward induction. *Mathematical Social Sciences*, **36**, 31–56.
Tan, T. & Werlang, S. (1988). The bayesian foundation of solution concepts of games. *Journal of Economic Theory*, **45**, 370–391.

6. Appendix

The following preliminary result is required in the proofs of Propositions 1 and 3; it may also be of independent interest.

LEMMA 1: *Fix an independent CPS $\delta_i \in I\Delta(S)$, a player $j \in N$ and a decreasing sequence $(A_j^k)_{k=0}^K$ of subsets of S_j such that $A_j^0 = S_j$ (assume $K > 0$). The following propositions are equivalent:*

(1) For every $s_j, s_j' \in S_j$: if $s_j \in A_j^m \setminus A_j^{m'}$, $s_j' \in A_j^{m'}$ and $K \geq m' > m$, then $\delta_{ij}(s_j' | \{s_j, s_j'\}) = 1$.
(2) For every $B_j, B_j' \in \mathcal{S}_j$: if $B_j \subseteq A_j^m \setminus A_j^{m'}$, $B_j' \subseteq A_j^{m'}$ and $K \geq m' > m$, then $\delta_{ij}(B_j' | B_j \cup B_j') = 1$.
(3) For every $B_j \subseteq S_j$ and $m = 0, 1, \ldots, K$, $A_j^m \cap B_j \neq \emptyset \Rightarrow \delta_{ij}(A_j^m | B_j) = 1$.

PROOF: (2) $\Rightarrow$ (1) is obvious. For the converse, assume that (1) holds. Then (2) holds if B_j, B_j' are singletons. By induction on the cardinality of these sets, consider first $\delta_{ij}(B_j | B_j \cup B_j' \cup b_j)$ with $b_j \in A_j^m \setminus A_j^{m'}$ and $b_j \notin B_j$.† Then, by Bayes' rule (Axiom 3) and the inductive hypothesis,

$$\begin{aligned}\delta_{ij}(B_j' | B_j \cup b_j \cup B_j') &= \delta_{ij}(B_j' | B_j \cup B_j') \delta_{ij}(B_j \cup B_j' | B_j \cup b_j \cup B_j') \\ &= \delta_{ij}(B_j \cup B_j' | B_j \cup b_j \cup B_j')\end{aligned}$$

so $\delta_{ij}(B_j | B_j \cup b_j \cup B_j') = 0$ by additivity. Similarly, $\delta_{ij}(b_j | B_j \cup b_j \cup B_j') = 0$, so $\delta_{ij}(B_J' | B_J \cup b_j \cup B_j') = 1$. Next, if $b_j' \in A_j^{m'}$ and $b_j' \notin B_j'$, by Bayes' rule and the inductive hypothesis,

$$\begin{aligned}\delta_{ij}(B_j' | B_j \cup B_j' \cup b_j') &= \delta_{ij}(B_j' | B_j \cup B_j') \delta_{ij}(B_j \cup B_j' | B_j \cup B_j' \cup b_j') \\ &= \delta_{ij}(B_j \cup B_j' | B_j \cup B_j' \cup b_j')\end{aligned}$$

so $\delta_{ij}(B_j | B_j \cup B_j' \cup b_j') = 0$ by additivity and $\delta_{ij}(B_j' \cup b_j' | B_j \cup B_j' \cup b_j') = 1$ as needed.

For (2) $\Rightarrow$ (3), let $\tilde{m} = \max\{m \in (0, \ldots, K) : B_j \cap A_j^m \neq \emptyset\}$. Since the sets A_j^m are nested, it is enough to show that $\delta_{ij}(B_j \cap A_j^{\tilde{m}} | B_j) = 1$. But this follows immediately from (2) by noting the decomposition

† Recall that we do not distinguish between a singleton and its unique element whenever the meaning is clear from the context.

$B_j = (B_j \cap A_j^{\tilde{m}}) \cup (B_j \setminus A_j^{\tilde{m}})$ and observing that $B_j \setminus A_j^{\tilde{m}} \subset A_j^m$ for some $m < \tilde{m}$.

Finally, to show (3) $\Rightarrow$ (2), let s_j, s_j' be as in the statement of the lemma and define $B_j = \{s_j, s_j'\}$. Then (3) implies that $1 = \delta_{ij}(A_j^{m'}|B_j) = \delta_{ij}(A_j^{m'} \cap B_j|B_j) = \delta_{ij}(s_j'|\{s_j, s_j'\})$.

PROOF OF PROPOSITION 1: The statement holds by definition for $k = 0$. Assume now that it is true for $k = 0, 1, \ldots, m$. Fix a strategy $s_i \in S_i^{m+1}$ and an independent CPS δ_i satisfying conditions (1) and (2) of Definition 2. By Lemma 1, we conclude that, for every $j \neq i$, $0 \leq k \leq m$ and $B_j \in \mathcal{S}_j$, $B_j \cap S_j^k \neq \emptyset \Rightarrow \delta_{ij}(S_j^k|B_j) = 1$. Since the induction hypothesis implies that $S_j^m \subseteq W_j^m$, and clearly $S_j \cap S_j^m \neq \emptyset$, we conclude that $\delta_{ij}(W_j^m|S_j) \geq \delta_{ij}(S_j^m|S_j) = 1$. Since δ_i satisfies the independence property, $\delta_i(\prod_{j \in N} W_j^m|S) = 1$: hence $s_i \in W_i^{m+1}$, as required.

The next ancillary result extends Lemma 5.1 in Battigalli and Siniscalchi (1997).

LEMMA 2: *Fix a player $i \in N$ and for every opponent $j \neq i$ a function $\tau_j : S_j \to T_j$. Then for every independent first order CPS $\delta_i \in I\Delta^{\mathcal{S}}(S)$ there is a unique type $t_i \in T_i$ such that $g_i(t_i) \in I\Delta^{\mathcal{S}}(\prod_{j=1}^n S_j \times T_j)$ and, for all $B \in \mathcal{S}$, $g_{i,B}(t_i)$ has finite support and satisfies:*

$$\forall s \in S, g_{i,B}(t_i)((s_i, t_i), (s_j, \tau_j(s_j))_{j \neq i}) = \delta_i(s|B).$$

PROOF: The existence and uniqueness of a type t_i satisfying the equalities above (so that $\delta_i = \delta_i(t_i)$) is proved in the cited source (the assumption that the associated function $\overline{g}_i$ is a homeomorphism is crucial). We only have to show that the type satisfies epistemic independence.

For any $J \subseteq N$ and $E_J \in \mathcal{E}_J$, let E_J^S be the subset of S_J corresponding to E_J given the functions τ_j ($j \in J$), that is,

$$E_J^S = \left\{ s_J \in S_J : [s_j, \tau_j(s_j)]_{j \in J} \in E_J \right\}.$$

Note that, since each function τ_j only depends on s_j, for any pair of subsets $E_J \in \mathcal{E}_J$, $F_{N \setminus J} \in \mathcal{E}_{N \setminus J}$, $(E_J \times F_{N \setminus J})^S = E_J^S \times F_J^S$. Now let $\tau_i(s_i) \equiv t_i$ and $\mu_i = g_i(t_i)$ for notational convenience. Fix $\emptyset \neq J \subset N$, $A_J \in \mathcal{E}_J$, $B_J \in \mathcal{C}_J(\mathcal{S}_J)$, $C_{N \setminus J} \in \mathcal{C}_{N \setminus J}(\mathcal{S}_{N \setminus J})$. Then the relation between t_i and δ_i and independence of δ_i imply

$$\mu_i(A_J \times C_{N \setminus J}|B_J \times C_{N \setminus J}) = \delta_i\left((A_J \times C_{N \setminus J})^S|(B_J \times C_{N \setminus J})^S\right) =$$

$$\delta_i\left(A_J^S \times C_{N \setminus J}^S|B_J^S \times C_{N \setminus J}^S\right) = \delta_i\left(A_J^S \times S_{N \setminus J}|B_J^S \times S_{N \setminus J}\right) =$$

$$\mu_i\left(A_J \times \left(\prod_{k\in N\setminus J} S_k \times T_k\right) | B_J \times \left(\prod_{k\in N\setminus J} S_k \times T_k\right)\right) = \mu_{iJ}(A_J|B_J).$$

Therefore μ_i is independent.

PROOF OF PROPOSITION 2: (1) We proceed by induction on k. At each step k, the proof of the "if" part of the statement is relatively straightforward; for the "only if" part, we show how to construct, for each $s_i \in W_i^{k+1}$, the type corresponding to s_i mentioned in the statement. This type is denoted $t_i^k(s_i)$. For brevity, we write $W^k = \prod_{j=1}^n W_j^k$. For every player $j \in N$ fix an arbitrary function $\tau_j : S_j \to T_j$.

(1.0) Suppose that $(s_j, t_j)_{j\in N} \in I \cap R$. Then for each $i \in N$, $s_i \in r_i(\delta_i(t_i))$ and $\delta_i(t_i)$ is independent, because $g_i(t_i)$ is independent. Therefore, $s \in W^1$.

Fix a player i and a strategy $s_i \in W_i^1$. By Definition 2 there is a corresponding CPS $\delta_i^0(s_i) \in I\Delta^{\mathcal{S}}(S)$ such that $s_i = r_i(\delta_i^0(s_i))$. Given the functions $\tau_j(\cdot)$ $(j \neq i)$, by Lemma 2 there is a corresponding type $t_i^0(s_i)$ such that $g_i(t_i^0(s_i))$ is independent, $\delta_i^0(s_i) = \delta_i(t_i^0(s_i))$ and, therefore, $s_i = \rho_i(t_i)$. For $s_i \in S_i\setminus W_i^1$, let $t_i^0(s_i) = \tau_i(s_i)$. This way we construct n functions $t_j^0(\cdot) : S_j \to T_j$ $(j \in N)$ such that, for all $s \in S$,

$$s \in W^1 \Rightarrow [s_j, t_j^0(s_j)]_{j\in N} \in I \cap R = \beta^0(I \cap R).$$

(1.k) Suppose, by way of induction, that the "if and only if" statement is true for $k-1$ and that we have constructed n functions $t_j^{k-1}(\cdot) : S_j \to T_j$ such that, for all $s \in S$,

$$s \in W^k \Rightarrow [s_j, t_j^{k-1}(s_j)]_{j\in N} \in \bigcap_{m=0}^{k-1} \beta^m(I \cap R).$$

Suppose that $(s_j, t_j)_{j\in N} \in \bigcap_{m=0}^{k} \beta^m(I \cap R)$. Then for each i, $s_i \in r_i(\delta_i(t_i))$, $\delta_i(t_i)$ is independent and $g_{i,S}(t_i)[\bigcap_{m=0}^{k-1} \beta^m(I \cap R)] = 1$. By the inductive hypothesis W^k is the projection of $\bigcap_{m=0}^{k-1} \beta^m(I \cap R)$ on S. Thus, the latter equality implies that $\delta_i(t_i)(W^k|S) = 1$. Therefore, $s \in W^{k+1}$.

Fix $i \in N$ and $s_i \in W_i^{k+1}$. By Definition 2 there is a corresponding CPS $\delta_i^k(s_i) \in I\Delta^{\mathcal{S}}(S)$ such that $s_i = r_i(\delta_i^k(s_i))$ and $\delta_i^k(s_i)(W^k|S) = 1$. Given the functions $t_j^{k-1}(\cdot)$ $(j \neq i)$, let $t_i^k(s_i) \in T_i$ be the independent type uniquely associated to $\delta_i^k(s_i)$ as in Lemma 2. This way we construct n functions $t_j^k(\cdot) : S_j \to T_j$ $(j \in N)$. We must show that, for all $s \in S$,

$$s \in W^{k+1} \Rightarrow [s_j, t_j^k(s_j)]_{j \in N} \in \bigcap_{m=0}^{k} \beta^m(I \cap R),$$

or equivalently

$$s \in W^{k+1} \Rightarrow [s_j, t_j^k(s_j)]_{j \in N} \in \beta^m(I \cap R), m = 0, \ldots, k.$$

The inductive thesis (2) follows from the construction of the $t_j^k(\cdot)$ $(j \in N)$ functions, the inductive hypothesis (1) and the introspection property. We provide a complete proof using a "subinductive" argument within this inductive step. For $m = 0$, (2) is clearly true.

Now suppose that (2) is true for $m \leq k-1$. We must show that $s \in W^{k+1}$ implies $(s_j, t_j^k(s_j)) \in \beta^{m+1}(I \cap R)$. Notice that, for each $i \in N$,

$$\beta_i^{m+1}(I \cap R) = \beta_i[\beta_i^m(I \cap R) \times \prod_{j \neq i} \beta_j^m(I \cap R)]$$

That is, for each player i, the inductive step entails both beliefs about *herself* and beliefs about her opponents. The former must be handled separately when $m = 0$.

Thus, fix i and $s_i \in W_i^{k+1}$. If $m = 0$, in order to show that (2) holds for $m + 1 = 1$ we invoke the assumption that every player is certain of her own type, and that every *rational* player is also certain of her strategy (see (1) in Definition 1) to conclude that:

$$[s_i, t_i^k(s_i)] \in \beta_i[(I_i \cap R_i]_i)$$

If $1 \leq m \leq k-1$, by the "subinductive" hypothesis $g_{i,S}[t_i^k(s_i)]$ $(\beta^{m-1}(I \cap R)) = 1$. Therefore, the introspection property (see remark 1) implies that $t_i^k(s_i)$ is certain of being certain of $\beta^{m-1}(I \cap R)$, that is,

$$g_{i,S}[t_i^k(s_i)]\left([\beta_i(\beta^{m-1}(I \cap R))]_i\right) = 1.$$

Notice that the introspection property concerns beliefs, and therefore does not entail any relevant restrictions when $m = 0$.

As for i's beliefs about the opponents, there is no need to distinguish cases. We have to show that for all $j \neq i$, $g_{i,S}(t_i^k(s_i))([\beta_j(\beta^{m-1}(I \cap R))]_j) = 1$. But this follows from the construction and the inductive hypothesis (2):

$$g_{i,S}(t_i^k(s_i))\left([\beta_j(\beta^{m-1}(I \cap R))]_j\right) =$$

$$\delta_{ij}^k(s_i)\left(\left\{s_j \in W_j^k : g_{j,S}\left[t_j^{k-1}(s_j)(\beta^{m-1}(I \cap R))\right] = 1 | S_j\right\}\right) = 1.$$

Therefore, $g_{i,S}[t_i^k(s_i)][\beta^m(I \cap R)] = 1$ as desired. Since this is true for all $i \in N$ and $s_i \in W_i^{k+1}$, then $s \in W^{k+1}$ implies $[s_i, t_i^k(s_i)]_{i \in N} \in \beta^{m+1}(I \cap S)$.

(2) The "if" part follows immediately from (1). For the "only if" part note that since S is finite and $W^{k+1} \subseteq W^k$ for all $k = 0, 1, \ldots$, there must be an integer K such that $W^K = \bigcap_{k \geq 0} W^k = W^{K+m}$ for all $m = 1, 2, \ldots$. Construct functions $t_j^K(\cdot) : S_j \to T_j$ $(j \in N)$ as in the proof of part (1). Then the same argument as above shows that for all $s \in W^K$ and $k = 0, 1, \ldots$, $(s_j, t_j^K(s_j)) \in \beta^k(I \cap R)$.

PROOF OF PROPOSITION 3: the following result implies part (1) immediately: *For all,* $k \geq 0$,

(a1) there are $2n$ functions $\delta_i^k : S_i \to I\Delta^{\mathcal{S}}((S_j)_{j \in N})$ and $t_i^k : S_i \to T_i$, $i \in N$, such that, for all $s \in S$ and $i \in N$, $\delta_i^k(s_i) = \delta_i(t_i^k(s_i))$, $g_i(t_i^k(s_i)) \in I\Delta^{\mathcal{S}}(\prod_{j=1}^n S_j \times T_j)$ and:

  (i) if $k \geq 1$, then for all $m = 0, \ldots, k-1$ and $i \in N$, $s_i \in S_i^m \setminus S_i^{m+1}$ implies $\delta_i^m(s_i) = \delta_i^k(s_i)$ and $t_i^m(s_i) = t_i^k(s_i)$,

  (ii) if $k \geq 1$, then for all $i \in N$, $s_i \in S_i^{k+1}$, $j \neq i$ and $s_j, s_j' \in S_j$: if $s_j \in S_j^m \setminus S_j^{m'}$, $s_j' \in S_j^{m'}$ and $k \geq m' > m$, then $\delta_{ij}^k(s_i)(s_j' | \{s_j, s_j'\}) = 1$.

  (iii) if, for all $i \in N$, $s_i \in S_i^{k+1}$, then $[s_i, t_i^k(s_i)] \in \gamma_i^k[(R_j \cap I_j)_{j \in N}]$ for all $i \in N$.

(a2) for all states $(s_j, t_j)_{j \in N}$, if $(s_j, t_j) \in \gamma_i^k[(R_j \cap I_j)_{j \in N}]$ for all $i \in N$ then $s_i \in S_i^{k+1}$ for all $i \in N$.

PROOF: $(k = 0)$ Fix $i \in N$. For $s_i \in S_i \setminus S_i^1$, choose $\delta_i^0(s_i) \in I\Delta^{\mathcal{S}}(S)$ arbitrarily; for $s_i \in S_i^1$, choose any element of $r_i^{-1}(s_i)$ which satisfies the independence condition (one must exist by the definition of S_i^1). By Lemma 2, an appropriate $t_i^0(s_i)$ can be found for every $s_i \in S_i$. Now the first two items in (a1) do not apply, while the third holds by construction. Finally, (a2) is obvious once one notices that, for any CPS $\mu_i \in I\Delta^{\mathcal{S}}(\prod_{j=1}^n S_j \times T_j)$, one may obtain an independent CPS $\delta_i \in I\Delta(S)$ by setting $\delta_i(A|B) = \mu_i[\mathcal{C}(A)|\mathcal{C}(B)]$ for every $A \subseteq S$ and $B \in \mathcal{S}$ (see Remark 2).

$(k = l)$ Suppose the statement above holds for $k = 0, \ldots, l-1$.

$(a1.l)$ For each $i \in N$ and $s_i \in S_i \setminus S_i^{l+1}$, let $\delta_i^l(s_i) = \delta_i^{l-1}(s_i)$ and $t_i^l(s_i) = t_i^{l-1}(s_i)$. Thus, the first claim in (a1) will hold regardless of how we complete the specification of δ_i^l and t_i^l on S_i^{l+1}.

By the definition, for each $s_i \in S_i^{l+1}$ we can find an independent CPS $\delta_i^l(s_i)$ on $(S_j)_{j \in N}$ such that $s_i = r_i[\delta_i^l(s_i)]$ and the second condition in (a1) is satisfied. Lemma 2 then yields a type $t_i^l(s_i)$ such that:

$$\forall s' \in S, B \in \mathcal{S}, \quad g_{i,B}[t_i^l(s_i)]\left([s_i', t_i, [t_j^{l-1}(s_j')]_{j \neq i}]\right) = \delta_i^l(s_i)(s'|B)$$

and

$$g_i(t_i^l(s_i)) \in I\Delta^{\mathcal{S}}\left(\prod_{j=1}^{n} S_j \times T_j\right).$$

Hence, $[s_i, t_i^l(s_i)] \in R_i \cap I_i$ for each $i \in N$. By Remark 3, to conclude the proof of the third point of (a1), we need to show that, for each $i \in N$, $m = 0, \ldots, k-1 = l-1$ and $j \neq i$, one has $(s_i, t_i^l(s_i)) \in \beta_{ij}^*(\gamma_j^m((R_z \cap I_z)_{z \in N}))$.

Thus fix $i \in N$. Suppose that, for some $m \in \{0, \ldots, l-1\}$ and $j \neq i$, $\gamma_j^m((R_z \cap I_z)_{z \in N}) \cap (B_j \times T_j) \neq \emptyset$ for some $B_j \in \mathcal{S}_j$. By the induction hypothesis, this implies that $S_j^{m+1} \cap B_j \neq \emptyset$.

Lemma 1 now implies that $\delta_{ij}^l(s_i)(S_j^{m+1}|B_j) = 1$. Also, observe that, for all $j \in N$, $z = m, \ldots, l-2$ and $s_j' \in S_j^{z+1} \setminus S_j^{z+2}$, the induction hypothesis implies that†

$$[s_j', t_j^{l-1}(s_j')] = [s_j', t_j^z(s_j')] \in \gamma_j^z[(R_w \cap I_w)_{w \in N}] \subseteq \gamma_j^m[(R_w \cap I_w)_{w \in N}]$$

and, for $s_j' \in S_j^l$,

$$[s_j', t_j^{l-1}(s_j')) \in \gamma_j^{l-1}[(R_w \cap I_w)_{w \in N}] \subseteq \gamma_j^m[(R_w \cap I_w)_{w \in N}]$$

Hence, for all $s_j' \in S_j^{m+1}$,

$$[s_j', t_j^{l-1}(s_j')] \in \gamma_j^m[(R_w \cap I_w)_{w \in N}].$$

By construction (and using independence),

$$g_{i,B_j \times S_{-j}}[t_i^l(s_i)]\left([\{(s_j', t_j^{l-1}(s_j'))\}]_j\right) = \delta_{ij}^l(s_i)(s_j'|B_j).$$

But then

$$\begin{aligned}
&g_{i,B_j \times S_{-j}}[t_i^l(s_i)]\left([\gamma_j^m((R_w \cap I_w)_{w \in N})]_j\right) \\
&\quad = g_{i,B_j \times S_{-j}}[t_i^l(s_i)]\left(\gamma_j^m[(R_w \cap I_w)_{w \in N}] \times \prod_{w \neq j}(S_w \times T_w)\right) \\
&\quad \geq \sum_{s_j' \in S_j^{m+1}} g_{i,B_j \times S_{-j}}[t_i^l(s_i)]\left(\{(s_j', t_j^{l-1}(s_j'))\} \times \prod_{w \neq j}(S_w \times T_w)\right) \\
&\quad = \sum_{s_j' \in S_j^{m+1}} \delta_{ij}^l(s_i)(s_j'|B_j) = \delta_{ij}^l(S_j^{m+1}|B_j) = 1
\end{aligned}$$

as required. This completes the proof of (a1).

† The proof is as follows: by the induction hypothesis, for any $z \in \{m, \ldots, l-2\}$, and for any $s \in S$ such that, for all $j \in N$, $s_j \in S_j^{z+1}$, we have $[s, t_j^z(s_j)] \in \gamma_j^z[(R_w \cap I_w)_{w \in N}]$. Now (again invoking the induction hypothesis) by construction $t_j^{l-1}(s_j') = t_j^{z+1}(s_j')$ if s_j' is eliminated at round $m+1$.

(a2.l). Suppose that $(s_i, t_i) \in \gamma_i^l[(R_j \cap I_j)_{j\in N}]$ for all $i \in N$. By monotonicity, $(s_i, t_i) \in \gamma_i^0((R_j \cap I_j)_{j\in N}) = R_i \cap I_i$, so (by the argument given in (a2.0)) for each $i \in N$ there exists an independent first-order CPS $\delta_i^l = \delta_i(t_i) \in I\Delta(\prod_{j\in N} S_j)$ such that $s_i = r_i(\delta_i^l)$. We must show that δ_i^l satisfies condition (2) of Definition 3.

We use Lemma 1. Consider any $j \neq i$ and $B_j \in \mathcal{S}_j$. Fix $m \in \{0, \ldots, l-1\}$, and suppose that $B_j \cap S_j^{m+1} \neq \emptyset$. The induction hypothesis implies that S_j^{m+1} is the projection of $\gamma_j^m(.)$ on S_j, so $(B_j \times T_j) \cap \gamma_j^m((R_w \cap I_w)_{w\in N}) \neq \emptyset$. By monotonicity, $(s_i, t_i) \in \gamma_j^m((R_w \cap I_w)_{w\in N})$, which implies that;

$$g_{i,B_j\times S_{-j}}(t_i)([\gamma_j^m((R_w \cap I_w)_{w\in N}))]_j) = 1$$

so that, again by the induction hypothesis,

$$\delta_{ij}^l(S_j^{m+1}|B_j) = [mrg_{S_j} g_{i,B_j\times S_{-j}}(t_i)](\pi_{S_j}\gamma_j^m((R_w \cap I_w)_{w\in N}))) = 1$$

as needed; π_{S_j} denotes projection on S_j. Since obviously $\delta_{ij}^l(S_j^0|B_j) = \delta_{ij}^l(S_j|B_j) = 1$, this completes the proof of the inductive step.

Part (2) follows once one notices that the sets $\gamma_i^l(.)$ are closed for every $i \in N$ and $l \geq 0$, that they form a nested sequence for each $i \in N$, and that each $S_i \times T_i$ is compact. This ensures that the infinite intersections appearing in the statement are non-empty. For details see the appendix of Battigalli and Siniscalchi (1997).

PROOF OF PROPOSITION 4: Let X^k denote the set of decision nodes x such that the longest path from x to a terminal node following x has k edges, and let $\iota(x)$ denote the player moving at x. Finally, let s^* be the unique subgame perfect equilibrium. We show by induction on k that if $x \in X^k$ and $s_{\iota(x)} \in W_{\iota(x)}^k$, then $s_{\iota(x)}(x) = s_{\iota(x)}^*(x)$. The statement is clearly true for $k = 1$. Suppose it is true for $k = 1, \ldots, m$. Let $x \in X^{m+1}$, $\iota(x) = i$, $s_i \in W_i^{m+1}$. Then there is an independent $\delta_i \in I\Delta^{\mathcal{S}}(S)$ such that $s_i \in r_i(\delta_i)$ and—for each j—$\delta_{ij}(W_j^m|S_j) = 1$. Let J be the set of players moving at nodes following x. By perfect information $S(x) = \prod_{j\in N} S_j(x)$. Since no player moves more than once, for all $j \in J \cup \{i\}$, $S_j(x) = S_j$. Therefore $\delta_{ij}[W_j^m|S_j(x)] = 1$ for all $j \in J$. Hence the inductive hypothesis implies that $\delta_i[\cdot|S(x)]$ assigns positive probability only to strategy profiles choosing the subgame perfect action at each node following x. By assumption, x is consistent with s_i, which must be a best response to $\delta_{i,-i}[\cdot|S(x)]$ at x. Since there are no relevant ties and i expects the subgame perfect path after each action, s_i must select the subgame perfect action $s_i^*(x)$.