



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Games and Economic Behavior 53 (2005) 15–42

GAMES and
Economic
Behavior

www.elsevier.com/locate/geb

Sequential and quasi-perfect rationalizability in extensive games

Geir B. Asheim^a, Andrés Perea^{b,*}

^a *Department of Economics, University of Oslo, PO Box 1095, Blindern, 0317 Oslo, Norway*

^b *Department of Quantitative Economics, Maastricht University, PO Box 616,
6200 MD Maastricht, The Netherlands*

Received 19 September 2002

Available online 12 October 2004

Abstract

Within an epistemic model for two-player extensive games, we formalize the event that each player believes that his opponent chooses rationally at all information sets. Letting this event be common certain belief yields the concept of *sequential rationalizability*. Adding preference for cautious behavior to this event likewise yields the concept of *quasi-perfect rationalizability*. These concepts are shown to (a) imply backward induction in generic perfect information games, and (b) be non-equilibrium analogues to sequential and quasi-perfect equilibrium, leading to epistemic characterizations of the latter concepts. Conditional beliefs are described by the novel concept of a system of *conditional lexicographic probabilities*.

© 2004 Elsevier Inc. All rights reserved.

JEL classification: C72

Keywords: Extensive games; Rationalizability

* Corresponding author.

E-mail addresses: g.b.asheim@econ.uio.no (G.B. Asheim), a.perea@ke.unimaas.nl (A. Perea).

1. Introduction

One of the major problems in the theory of extensive games is the following: how should a player react when he finds himself at an information set that contradicts his previous belief about the opponent's strategy choice? Different approaches have been proposed to this problem. Ben-Porath (1997) and Reny (1992) have formulated rationalizability and equilibrium notions based on *weak sequentiality*, in which a player is allowed to believe, in this situation, that his opponent will no longer choose rationally. Battigalli and Siniscalchi (2002) have shown that Pearce's (1984) *extensive form rationalizability* can be characterized by assuming that a player, in such a situation, should look for the highest degree of "strategic sophistication" that is compatible with the event of reaching this information set, and stick to this degree until it is contradicted later on in the game. Perea (2002, 2003) suggests that the player, in such a situation, may revise his conjecture about the opponent's utility function in order to rationalize her "surprising" move, while maintaining common belief of rational choice at all information sets. The most prominent position, however, is that the player should still believe that his opponent will choose rationally in the remainder of the game; this underlies concepts that promote backward induction. We are concerned with such concepts in the present paper.

We define *sequential rationalizability* by imposing common 'certain belief' of the event that each player believes that the opponent chooses rationally at all her information sets.¹ We define *quasi-perfect rationalizability* by imposing common 'certain belief' of the event that each player has preference for cautious behavior (i.e., at every information set, one strategy is preferred to another if the former weakly dominates the latter) and believes that the opponent chooses rationally at all her information sets. Since these are non-equilibrium concepts, each player need not be certain of the beliefs that the opponent has about the player's own action choice. However, by *assuming* that each player is certain of the beliefs that the opponent has about the player's own action choice, we obtain epistemic characterizations of the corresponding equilibrium concepts: *sequential* and *quasi-perfect equilibrium*. When applied to generic games with perfect information, both sequential and quasi-perfect rationalizability yield the backward induction procedure. To avoid the issue of whether (and if so, how) each player's beliefs about the action choice of his opponents are stochastically independent, all analysis is limited to two-player games.

For the above mentioned definitions and characterizations, we must describe what a player believes both conditional on reaching his own information sets (to evaluate his rationality) and conditional on his opponent reaching her information sets (to determine his beliefs about her choices). In other words, we must specify a system of conditional beliefs for each player. There are various ways to do so. One possibility is a *conditional probability system* (CPS) where each conditional belief is a subjective probability distribution.² This is sufficient to model sequentiality in the current context. Another possibility, which is sufficient to model quasi-perfectness in the current context, is to apply a single sequence of

¹ 'Certain belief,' which is the operator we will use for the interactive epistemology, will be defined in Section 2.4. An event is said to be 'certainly believed' if the complement is deemed subjectively impossible.

² We use Myerson's (1986) terminology. In philosophical literature, related concepts are called Popper measures. For an overview over relevant literature and analysis, see Halpern (2003).

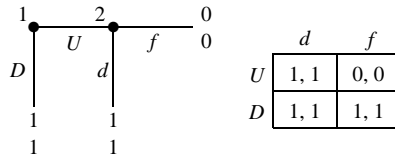


Fig. 1. Γ_1 and its strategic form.

subjective probability distributions—a so-called *lexicographic probability system* (Blume et al., 1991a, LPS)—and derive the conditional beliefs as the conditionals of such an LPS. Since each conditional LPS is found by constructing a new sequence which includes the well-defined conditional probability distributions of the original sequence (see footnote 3), each conditional belief is itself an LPS.

However, quasi-perfectness cannot always be modelled by a CPS since the modelling of preference for cautious behavior may require lexicographic probabilities. To see this, consider Γ_1 of Fig. 1.

In this game, if player 1 believes that player 2 chooses rationally, then player 1 must assign probability one to player 2 choosing *d*. Hence, if each (conditional) belief is associated with a probability distribution—as is the case with the concept of a CPS—and player 1 believes that his opponent chooses rationally, then player 1 is indifferent between his two strategies. This is inconsistent with quasi-perfectness, which requires players to have preference for cautious behavior, meaning that player 1 in Γ_1 prefers *D* to *U*.

Moreover, sequentiality cannot always be modelled by means of conditionals of a single LPS since preference for cautious behavior is induced. To see this, consider a modified version of Γ_1 where an additional subgame is substituted for the (0, 0)-payoff, with all payoffs in that subgame being smaller than 1. If player 1’s conditional beliefs over strategies for player 2 is derived from a single LPS, then a well-defined belief conditional on reaching the added subgame entails that player 1 deems possible the event that player 2 chooses *f*, and hence, player 1 prefers *D* to *U*. This is inconsistent with sequentiality, under which *U* is a rational choice.

We therefore introduce a new way of describing a system of conditional beliefs, called a *system of conditional lexicographic probabilities* (SCLP). In contrast to a CPS, an SCLP may induce conditional beliefs that are represented by LPSs rather than subjective probability distributions. In contrast to the system of conditionals derived from a single LPS, an SCLP need not include all levels in the sequence of the original LPS when determining conditional beliefs.

It is our aim to model sequential rationalizability (and equilibrium) and quasi-perfect rationalizability (and equilibrium) within the same epistemic model. By embedding the notion of an SCLP in an epistemic model with a set of epistemic types for each player, we will be able to model quasi-perfectness as a special case of sequentiality. For each type t_i of any player i , t_i is described by an SCLP, inducing a behavior strategy for each opponent type t_j that is deemed subjectively possible by t_i . The event that “player i believes that the opponent j chooses rationally at each information set” can then be defined as the event where player i is of a type t_i that, for each subjectively possible opponent type t_j , induces a behavioral strategy which is sequentially rational given t_j ’s own SCLP.

An SCLP ensures well-defined conditional beliefs representing nontrivial conditional preferences, while allowing for flexibility w.r.t. whether to assume preference for cautious behavior. Preference for cautious behavior, as needed for quasi-perfect rationalizability, is obtained by imposing the following additional requirement on t_i 's SCLP for each conditioning event: if an opponent strategy-type pair (s_j, t_j) is compatible with the event and t_j is deemed subjectively possible by t_i , then (s_j, t_j) is in the support of type t_i 's conditional belief.

The concept of sequential rationalizability is related to various other concepts proposed in the literature. Already in Bernheim (1984) there are suggestions concerning how to define non-equilibrium concepts that involve rational choice at all information sets. By requiring rationalizability in every subgame, Bernheim defines the concept of *subgame rationalizability*—which coincides with our definition of sequential rationalizability for games of almost perfect information—but no epistemic characterization is offered. Bernheim (1984, p. 1022) claims that it is possible to define a concept of sequential rationalizability, but does not indicate how this can be done. After related work by Greenberg (1996), sequential rationalizability was finally defined by Dekel et al. (1999, 2002), whose concept coincides with ours in our two-player setting. Our definition of quasi-perfect rationalizability is new. Dekel et al. (1999) and Greenberg et al. (2003) consider also extensive game concepts that lie between equilibrium and rationalizability; such concepts will not be considered here.

Our paper is organized as follows. In Section 2 we define the concept of an SCLP and introduce the epistemic model that will be used throughout the paper. In Section 3 we present our epistemic characterizations of sequential and quasi-perfect equilibria and define the concepts of sequential and quasi-perfect rationalizability. In Section 4 we investigate the relationship to other rationalizability concepts, while in Section 5 we show how sequential and quasi-perfect rationalizability promote backward induction. In Section 6 we discuss the restriction to two-player games. A representation result for SCLP is established in Appendix A. Proofs of the main characterization results are contained in Appendix B, while proofs of others results are available on request.

2. Players as decision makers

In this section, we introduce some definitions and notation in order to model the players in an extensive game as decision-makers under uncertainty.

2.1. A system of conditional lexicographic probabilities

Consider a decision-maker under uncertainty, and let F be a finite set of states. The decision-maker is uncertain about what state in F will be realized. Let F^* ($\subseteq F$) be the nonempty subset of states that the decision maker deems subjectively possible. Write $\mathcal{F}^* := \{E \subseteq F \mid E \cap F^* \neq \emptyset\}$. Let Z be a finite set of outcomes. For any $E \in \mathcal{F}^*$, the decision-maker is endowed with complete and transitive conditional preferences over all functions that to each element of E assigns an objective randomization on Z . Any such function is called an Anscombe and Aumann (1963) *act* on E .

Refer to the collection of conditional preferences for all $E \in \mathcal{F}^*$ as a *system of conditional preferences*. We show in Proposition A.1 of Appendix A how such a system of conditional preferences can be represented by our novel notion of an SCLP (cf. Section 1). To introduce this notion formally, we need some preliminaries.

Let $v : Z \rightarrow \mathbb{R}$ be a vNM utility function, and abuse notation slightly by writing $v(p) = \sum_{z \in Z} p(z)v(z)$ whenever $p \in \Delta(Z)$ is an objective randomization.

A *lexicographic probability system* (LPS) consists of L levels of subjective probability distributions: if $L \geq 1$ and, $\forall \ell \in \{1, \dots, L\}$, $\mu_\ell \in \Delta(F)$, then $\lambda = (\mu_1, \dots, \mu_L)$ is an LPS on F . Let $\mathbf{L}\Delta(F)$ denote the set of LPSs on F . Write $\text{supp } \lambda := \bigcup_{\ell=1}^L \text{supp } \mu_\ell$. If $\text{supp } \lambda \cap E \neq \emptyset$, denote by $\lambda|_E = (\mu'_1, \dots, \mu'_{L|E})$ the conditional of λ on E .³ Write, for any $\ell \in \{1, \dots, L\}$, $\lambda_\ell := (\mu_1, \dots, \mu_\ell)$.

Definition 1. A *system of conditional lexicographic probabilities* (SCLP) (λ, ℓ) on F with support F^* consists of

- an LPS $\lambda = (\mu_1, \dots, \mu_L) \in \mathbf{L}\Delta(F)$ satisfying $\text{supp } \lambda = F^*$, and
- a function $\ell : \mathcal{F}^* \rightarrow \{1, \dots, L\}$ satisfying
 - (i) $\text{supp } \lambda_{\ell(E)} \cap E \neq \emptyset$,
 - (ii) $\ell(D) \geq \ell(E)$ whenever $\emptyset \neq D \subseteq E$, and
 - (iii) $\ell(\{e\}) \geq \ell$ whenever $e \in \text{supp } \mu_\ell$.

The interpretation is that the conditional belief on E is given by the conditional on E of the LPS $\lambda_{\ell(E)}$, $\lambda_{\ell(E)}|_E = (\mu'_1, \dots, \mu'_{\ell(E)|E})$. To determine preference between acts conditional on E , first calculate expected utilities by means of the top level probability distribution, μ'_1 , and then, if necessary, use the lower level probability distributions, $\mu'_2, \dots, \mu'_{\ell(E)|E}$, lexicographically to resolve ties. The function ℓ thus determines, for every event E , the number of levels of the original LPS λ that can be used, provided that they intersect with E , to resolve ties between acts conditional on E . Condition (i) ensures well-defined conditional beliefs that represent nontrivial conditional preferences. Condition (ii) means that the system of conditional preferences is dynamically consistent, in the sense that strict preference between two acts would always be maintained if new information, ruling out states at which the two acts lead to the same outcomes, became available. To motivate condition (iii), note that if $e \in \text{supp } \mu_\ell$ and $\ell(\{e\}) < \ell$, then it follows from condition (ii) that μ_ℓ could as well ignore e without changing the conditional beliefs.

A *full support SCLP* (i.e., an SCLP where $F^* = F$) combines the structural implication of a full support LPS—namely that conditional preferences are nontrivial—with flexibility w.r.t. whether to assume the behavioral implication of any conditional of such an LPS—namely that the conditional LPS’s full support induces preference for cautious behavior. A full support SCLP is a generalization of both

³ I.e., $\forall \ell \in \{1, \dots, L|E\}$, $\mu'_\ell(\cdot) = \mu_{k_\ell}(\cdot|E)$, where the indices k_ℓ are given by $k_0 = 0$, $k_\ell = \min\{k | \mu_k(E) > 0$ and $k > k_{\ell-1}\}$ for $\ell > 0$, and $\{k | \mu_k(E) > 0$ and $k > k_{L|E}\} = \emptyset$, and where $\mu_{k_\ell}(\cdot|E)$ is given by the usual definition of conditional probabilities; cf. Blume et al. (1991a, Definition 4.2).

- (1) Conditional beliefs described by a single full support LPS $\lambda = (\mu_1, \dots, \mu_L)$: let, for all $E \in \mathcal{F}^*$, $\ell(E) = L$. Then the conditional belief on E is described by the conditional of λ on E , $\lambda|_E$.
- (2) Conditional beliefs described by a CPS: let, for all $E \in \mathcal{F}^*$, $\ell(E) = \min\{\ell \mid \text{supp } \lambda_\ell \cap E \neq \emptyset\}$. Then, it follows from conditions (ii) and (iii) of Definition 1 that the full support LPS $\lambda = (\mu_1, \dots, \mu_L)$ has non-overlapping supports (i.e., λ is a *lexicographic conditional probability system* in the terminology of Blume et al., 1991a, Definition 5.2) and the conditional belief on E is described by the top level probability distribution of the conditional of λ on E . This corresponds to the isomorphism between CPS and lexicographic conditional probability system noted by Blume et al. (1991a, p. 72) and discussed by Halpern (2003).

However, a full support SCLP may describe a system of conditional beliefs that is not covered by these special cases. The following is a simple example: Let $F^* = F = \{d, e, f\}$ and $\lambda = (\mu_1, \mu_2)$, where $\mu_1(d) = 1/2$, $\mu_1(e) = 1/2$, and $\mu_2(f) = 1$. If $\ell(F) = 1$ and $\ell(E) = 2$ for any other non-empty subset E , then the resulting SCLP falls outside cases (1) and (2).

2.2. An extensive game

Consider a finite *extensive game form* with two players. Assume that there are no chance moves, and that the extensive game form satisfies perfect recall. Denote by H_i the collection of information sets controlled by player i . For every information set $h \in H_i$, let $A(h)$ be the set of actions available at h . A *pure strategy* for player i is a function s_i which assigns to every information set $h \in H_i$ some action $s_i(h) \in A(h)$. Denote by S_i the set of pure strategies for player i , where there, in the subsequent analysis, is no need to differentiate between pure strategies in S_i that differ only at non-reachable information sets. Write $S = S_1 \times S_2$ and denote by Z the set of terminal nodes (or outcomes). Let $z: S \rightarrow Z$ map strategy profiles into terminal nodes, and refer to (S_1, S_2, z) as the associated *strategic game form*.

Let, for each i , $v_i: Z \rightarrow \mathbb{R}$ be a vNM utility function that assigns a payoff to any outcome. Then the pair of the extensive game form and the vNM utility functions (v_1, v_2) constitutes a finite *extensive game* Γ . Let $G = (S_1, S_2, u_1, u_2)$ be the associated finite *strategic game*, where for each i , the vNM utility function $u_i: S \rightarrow \mathbb{R}$ is defined by $u_i = v_i \circ z$ (i.e., $u_i(s) = v_i(z(s))$ for any $s = (s_1, s_2) \in S$). Assume that, for each i , there exist $s, s' \in S$ such that $u_i(s) > u_i(s')$.

For any $h \in H_1 \cup H_2$, let $S_i(h)$ be the set of strategies s_i for which there is some strategy s_j such that (s_i, s_j) reaches h . For any h and any node $x \in h$, denote by $S(x) = S_1(x) \times S_2(x)$ the set of pure strategy profiles for which x is reached, and write $S(h) := \bigcup_{x \in h} S(x)$. By perfect recall, it holds that $S(h) = S_1(h) \times S_2(h)$ for all information sets h . For any $h, h' \in H_i$, h (weakly) precedes h' if and only if $S(h) \supseteq S(h')$. For any $h \in H_i$ and $a \in A(h)$, write $S_i(h, a) := \{s_i \in S_i(h) \mid s_i(h) = a\}$.

A *behavior strategy* for player i is a function σ_i that assigns to every $h \in H_i$ some randomization $\sigma_i(h) \in \Delta(A(h))$ on the set of available actions. If $h \in H_i$, denote by $\sigma_i|_h$ the behavior strategy with the following properties:

- (1) at player i information sets preceding h , $\sigma_i|_h$ determines with probability one the unique action leading to h , and
- (2) at all other player i information sets, $\sigma_i|_h$ coincides with σ_i .

Say that σ_i is *outcome-equivalent* to a mixed strategy p_i ($\in \Delta(S_i)$) if, for any $s_j \in S_j$, σ_i and p_i induce the same probability distribution over terminal nodes. For any $h \in H_i$, $\sigma_i|_h$ is outcome-equivalent to some $p_i \in \Delta(S_i(h))$.

2.3. Types

When an extensive game form is turned into a decision problem for each player, the uncertainty faced by a player concerns the action choice of his opponent at each of her information sets, the belief of his opponent about the player’s own action choice at each of his information sets, and so on. A *type* of a player in an extensive game form corresponds to a vNM utility function and a belief about the action choice of his opponent at each of her information sets, a belief about the belief of his opponent about the player’s own action choice at each of his information sets, and so on.

An implicit model with a finite set of type profiles, $T = T_1 \times T_2$, describes such hierarchies of beliefs. Each type $t_i \in T_i$ of any player i corresponds to i ’s vNM utility function v_i and a system of conditional beliefs on $S_j \times T_j$. For given $t_j \in T_j$ (i.e., for given belief of j about i ’s action choice at each $h \in H_i$, belief of j about i ’s belief about j ’s action choice at each $h \in H_j$, and so on), t_i ’s belief about j ’s action at some $h' \in H_j$ can be derived from his conditional belief on $S_j(h') \times \{t_j\}$ since the set of actions available at h' , $A(h')$, corresponds to a partition of $S_j(h')$.

For each $t_i \in T_i$, let $T_j^{t_i} (\subseteq T_j)$ be the non-empty set of opponent types that t_i deems subjectively possible. Also, assume that, for all $s_j \in S_j$ and $t_j \in T_j^{t_i}$, t_i deems (s_j, t_j) subjectively possible. This means that conditional beliefs are well-defined for an event $E_j (\subseteq S_j \times T_j)$ if and only if $E_j \cap (S_j \times T_j^{t_i}) \neq \emptyset$. Note that $\{S_j(h) \times T_j \mid h \in H_i\}$ is the set of events that are objectively observable by t_i . Hence, conditional beliefs are always well-defined for such events since, for any $h \in H_i$,

$$(S_j(h) \times T_j) \cap (S_j \times T_j^{t_i}) = (S_j(h) \times T_j^{t_i}) \neq \emptyset.$$

By describing the system of conditional beliefs by means of an SCLP, our construction, formulated within the strategic game form, can be summarized as follows.

Definition 2. For given vNM utility functions (v_1, v_2) on $\Delta(Z)$, an *epistemic model* for a strategic game form (S_1, S_2, z) consists of

- for each player i , a finite set of types T_i , and
- for each type t_i of any player i , and SCLP $(\lambda^{t_i}, \ell^{t_i})$ on $S_j \times T_j$ with support $S_j \times T_j^{t_i}$.

To illustrate this use of our notion of an SCLP, consider again the game of Fig. 1. Suppose that $T = \{t_1\} \times \{t_2\}$, and let $\lambda^{t_1} = (\mu_1^{t_1}, \mu_2^{t_1}) \in \mathbf{L}\Delta(S_2 \times \{t_2\})$ be such that $\mu_1^{t_1}$ assigns probability one to (d, t_2) and $\mu_2^{t_1}$ assigns positive probability to (f, t_2) . If $\ell^{t_1}(S_2 \times \{t_2\}) = 1$, then the SCLP corresponds to a CPS, while if $\ell^{t_1}(S_2 \times \{t_2\}) = 2$, then all conditional beliefs

are conditionals of λ^{t_1} . Condition (i) of Definition 1 requires that $\ell^{t_1}(\{(f, t_2)\}) = 2$, while condition (ii) implies that $\ell^{t_1}(\{(d, t_2)\}) \geq \ell^{t_1}(S_2 \times \{t_2\})$. Hence, the function ℓ^{t_1} yields flexibility w.r.t. whether to assume preference for cautious behavior, while ensuring that all conditional beliefs are well-defined.

2.4. Certain belief

In Definition 2 we allow for the possibility that each player deems some opponent types subjectively impossible, corresponding to an SCLP that does not have full support along the type dimension. Therefore, the epistemic operator ‘certain belief’ (meaning that the complement is subjectively impossible) can be derived from the epistemic model and defined for events that are subsets of $T_1 \times T_2$. For any $A \subseteq T_1 \times T_2$, say that at (t_1, t_2) player i *certainly believes* the event A if $(t_1, t_2) \in K_i A$, where

$$K_i A := \{(t_1, t_2) \in T_1 \times T_2 \mid \{t_i\} \times T_j^{t_i} \subseteq A\}.$$

Say that there is *mutual certain belief* of $A \subseteq T_1 \times T_2$ at (t_1, t_2) if $(t_1, t_2) \in KA$, where $KA := K_1 A \cap K_2 A$. Say that there is *common certain belief* of $A \subseteq T_1 \times T_2$ at (t_1, t_2) if $(t_1, t_2) \in CKA$, where $CKA := KA \cap KKA \cap KKA \cap \dots$.

2.5. Preferences over strategies

In an extensive game, player i makes decisions at his information sets. At every information set $h \in H_i$, the combination of i 's vNM utility function v_i and t_i 's conditional belief on $S_j(h) \times T_j$ determines complete and transitive preferences $\succsim_h^{t_i}$ on the set of acts from $S_j(h) \times T_j$ to Z . Since each strategy $s_i \in S_i(h)$ is a function that assigns $z(s_i, s_j)$ to any $(s_j, t_j) \in S_j(h) \times T_j$ and is thus an act from $S_j(h) \times T_j$ to Z , we have that $\succsim_h^{t_i}$ determines complete and transitive preferences on $S_i(h)$.

The *choice function* for type t_i of any player i is a function $C_i^{t_i}$ that assigns to every $h \in H_i$ t_i 's set of rational strategies:

$$C_i^{t_i}(h) := \{s_i \in S_i(h) \mid s_i \succsim_h^{t_i} s'_i \text{ for all } s'_i \in S_i(h)\}.$$

3. Sequential and quasi-perfect rationalizability

In this section, we use the concept of an SCLP to formalize the requirement that each player believes that the opponent chooses rationally at each of her information sets, given her preferences at these information sets. This enables us

- to characterize sequential (Kreps and Wilson, 1982) and quasi-perfect (van Damme, 1984) equilibrium,⁴ and

⁴ The concept of a quasi-perfect equilibrium differs from Selten's (1975) extensive form perfect equilibrium by the property that, at each information set, the player taking an action ignores the possibility of his own future mistakes.

- to define *sequential* and *quasi-perfect rationalizability* as non-equilibrium analogues to the concepts of Kreps and Wilson (1982) and van Damme (1984) in two-player extensive games.

3.1. Sequentiality

In our setting a behavior strategy is not an object of choice, but an expression of the system of beliefs of the other player. Say that the behavior strategy $\sigma_j^{t_i|t_j}$ is induced for t_j by t_i if $t_j \in T_j^{t_i}$ and, for all $h \in H_j$ and $a \in A(h)$,

$$\sigma_j^{t_i|t_j}(h)(a) := \frac{\mu_\ell^{t_i}(S_j(h, a), t_j)}{\mu_\ell^{t_i}(S_j(h), t_j)},$$

where ℓ is the first level ℓ of λ^{t_i} for which $\mu_\ell^{t_i}(S_j(h), t_j) > 0$, implying that $\mu_\ell^{t_i}$ restricted to $S_j(h) \times \{t_j\}$ is proportional to the top level probability distribution of the LPS that describes t_i 's conditional belief on $S_j(h) \times \{t_j\}$. Here, $\mu_\ell^{t_i}(S_j(h), t_j)$ is a short way to write $\mu_\ell^{t_i}(S_j(h) \times \{t_j\})$. Similarly for $\mu_\ell^{t_i}(S_j(h, a), t_j)$.

Say that the behavior strategy σ_i is *sequentially rational* for t_i if,

$$\forall h \in H_i, \quad \sigma_i|_h \text{ is outcome-equivalent to some mixed strategy in } \Delta(C_i^{t_i}(h)).$$

Define the event that player i is of a type that induces a sequentially rational behavior strategy for any opponent type that is deemed subjectively possible:

$$[isr_i] := \{(t_1, t_2) \in T_1 \times T_2 \mid \forall t'_j \in T_j^{t_i}, \sigma_j^{t_i|t'_j} \text{ is sequentially rational for } t'_j\}.$$

Write $[isr] := [isr_1] \cap [isr_2]$ for the event where both players are of such a type.

Note that the behavior strategy induced for t'_j by t_i specifies t_i 's *belief revision policy* about the behavior of t'_j , as it defines probability distributions *also* at player j information sets that are *unreachable* given t_i 's initial belief about t'_j 's behavior. Hence, if the true type profile (t_1, t_2) is in $[isr_i]$, then player i believes that each subjectively possible opponent type t'_j chooses rationally also at player j information sets that contradict t_i 's initial belief about the behavior of t'_j . The above observation explains why we can *characterize* a sequential equilibrium as a profile of induced behavior strategies at a type profile in $[isr]$ where there is mutual certain belief of the type profile (i.e., for each player, only the true opponent type is deemed subjectively possible).

Before doing so, we define sequential equilibrium. Player i 's beliefs over past opponent actions at i 's information sets is a function β_i that to any $h \in H_i$ assigns a probability distribution over the nodes in h . An assessment $(\sigma, \beta) = ((\sigma_1, \sigma_2), (\beta_1, \beta_2))$, consisting of a pair of behavior strategies and a pair of beliefs, is consistent if there is a sequence $(\sigma^n, \beta^n)_{n \in \mathbb{N}}$ of assessments converging to (σ, β) such that for every n , σ^n is completely mixed and β^n is induced by σ^n using Bayes' rule. If σ_i and σ_j are any behavior strategies for i and j , and β_i are the beliefs of i , then let, for each $h \in H_i$, $u_i(\sigma_i, \sigma_j; \beta_i)|_h$ denote i 's expected payoff conditional on h , given the belief $\beta_i(h)$, and given that future behavior is determined by σ_i and σ_j .

Definition 3. An assessment $(\sigma, \beta) = ((\sigma_1, \sigma_2), (\beta_1, \beta_2))$ is a *sequential equilibrium* if it is consistent and it satisfies for each i and every $h \in H_i$,

$$u_i(\sigma_i, \sigma_j; \beta_i)|_h = \max_{\sigma'_i} u_i(\sigma'_i, \sigma_j; \beta_i)|_h.$$

The characterization result can now be stated; it is proven in Appendix B.

Proposition 4. Consider a finite extensive two-player game Γ . A profile of behavior strategies $\sigma = (\sigma_1, \sigma_2)$ can be extended to a sequential equilibrium if and only if there exists an epistemic model with $(t_1, t_2) \in [isr]$ such that (1) there is mutual certain belief of $\{(t_1, t_2)\}$ at (t_1, t_2) , and (2) for each i , σ_i is induced for t_i by t_j .

For the “if” part, it is sufficient that there is mutual certain belief of the beliefs that each player has about the action choice of his opponent at each of her information sets. We do not need the stronger condition that (1) entails. Hence, higher order certain belief plays no role in the characterization, in line with the fundamental insights of Aumann and Brandenburger (1995).⁵

We next *define* the concept of sequentially rationalizable behavior strategies as induced behavior strategies under common certain belief of $[isr]$.

Definition 5. A behavior strategy σ_i for i is *sequentially rationalizable* in a finite extensive two-player game Γ if there exists an epistemic model with $(t_1, t_2) \in CK[isr]$ such that σ_i is induced for t_i by t_j .

It follows from Proposition 4 that a behavior strategy is sequentially rationalizable if it is part of a profile of behavior strategies that can be extended to a sequential equilibrium. Since a sequential equilibrium always exists, we obtain as an immediate consequence that sequentially rationalizable behavior strategies always exist.

3.2. Quasi-perfectness

Impose the additional requirement that for each type t_i of any player i the full LPS λ^{t_i} is used to form the conditional beliefs over opponent strategy-type pairs. Formally, let L be the number of levels in the LPS λ^{t_i} and define the event

$$[cau_i] := \{(t_1, t_2) \in T_1 \times T_2 \mid \ell^{t_i}(S_j \times T_j) = L\}.$$

Since ℓ^{t_i} is non-increasing w.r.t. set inclusion, $(t_1, t_2) \in [cau_i]$ implies that $\ell^{t_i}(E_j) = L$ for all subsets E_j of $S_j \times T_j$ with well-defined conditional beliefs. Due to the assumption that λ^{t_i} has full support on S_j (cf. Definition 2), $(t_1, t_2) \in [cau_i]$ means that t_i 's choice function never admits a weakly dominated strategy, thereby inducing preference for *cautious* behavior. Write $[cau] := [cau_1] \cap [cau_2]$.

We now *characterize* the concept of a quasi-perfect equilibrium as profiles of induced behavior strategies at a type profile in $[isr] \cap [cau]$ where there is mutual certain belief of

⁵ We are grateful to the referee for making this observation.

the type profile. To state the definition of quasi-perfect equilibrium, we need some preliminary definitions. Define the concepts of a behavior representation of a mixed strategy and the mixed representation of a behavior strategy in the standard way (cf., e.g., Myerson, 1991, p. 159). If a behavior strategy σ_j and a mixed strategy p_j are both completely mixed, and σ_j is a behavior representation of p_j or p_j is the mixed representation of σ_j , then, $\forall h \in H_j, \forall a \in A(h)$,

$$\sigma_j(h)(a) = \frac{p_j(S_j(h, a))}{p_j(S_j(h))}.$$

If p_j is a completely mixed strategy and $h \in H_i$, let $p_j|_h$ be defined by

$$p_j|_h(s_j) = \begin{cases} \frac{p_j(s_j)}{p_j(S_j(h))}, & \text{if } s_j \in S_j(h), \\ 0, & \text{otherwise.} \end{cases}$$

If σ_i is any behavior strategy for i and σ_j is a completely mixed behavior strategy for j , then abuse notation slightly by writing, for each $h \in H_i$,

$$u_i(\sigma_i, \sigma_j)|_h := u_i(p_i, p_j|_h),$$

where p_i is outcome-equivalent to $\sigma_i|_h$ and p_j is the mixed representation of σ_j .

Definition 6. A behavior strategy profile $\sigma = (\sigma_1, \sigma_2)$ is a *quasi-perfect equilibrium* if there is a sequence $(\sigma^n)_{n \in \mathbb{N}}$ of completely mixed behavior strategy profiles converging to σ such that for each i and every $n \in \mathbb{N}$ and $h \in H_i$,

$$u_i(\sigma_i, \sigma_j^n)|_h = \max_{\sigma'_i} u_i(\sigma'_i, \sigma_j^n)|_h.$$

The characterization result can now be stated; it is proven in Appendix B.

Proposition 7. Consider a finite extensive two-player game Γ . A profile of behavior strategies $\sigma = (\sigma_1, \sigma_2)$ is a *quasi-perfect equilibrium* if and only if there exists an epistemic model with $(t_1, t_2) \in [isr] \cap [cau]$ such that

- (1) there is mutual certain belief of $\{(t_1, t_2)\}$ at (t_1, t_2) , and
- (2) for each i , σ_i is induced for t_i by t_j .

As for Proposition 4, higher order certain belief plays no role in this characterization.

We next *define* the concept of quasi-perfectly rationalizable behavior strategies as induced behavior strategies under common certain belief of $[isr] \cap [cau]$.

Definition 8. A behavior strategy σ_i for i is *quasi-perfectly rationalizable* in a finite extensive two-player game Γ if there exists an epistemic model with $(t_1, t_2) \in CK([isr] \cap [cau])$ such that σ_i is induced for t_i by t_j .

It follows from Proposition 7 that a behavior strategy is quasi-perfectly rationalizable if it is part of a quasi-perfect equilibrium. Since a quasi-perfect equilibrium always exists, we

obtain as an immediate consequence that quasi-perfectly rationalizable behavior strategies always exist.

Propositions 4 and 7 imply the well-known result that every quasi-perfect equilibrium can be extended to a sequential equilibrium, while Definitions 5 and 8 imply that the set of quasi-perfectly rationalizable strategies is included in the set of sequentially rationalizable strategies. To illustrate that this inclusion can be strict, consider Γ_1 of Fig. 1 (cf. Section 2). Both concepts predict that player 2 plays d with probability one. However, only quasi-perfect rationalizability predicts that player 1 plays D with probability one. Preferring D to U amounts to preference for cautious behavior since by choosing D player 1 avoids the risk that player 2 may choose f .

4. Relation to other rationalizability concepts

In this section, we explore the relationship between sequential and quasi-perfect rationalizability, on the one hand, and the concepts of rationalizability, permissibility, weak sequential rationalizability, extensive form rationalizability, and proper rationalizability, on the other hand. We have observed in the previous section that sequential and quasi-perfect rationalizability may be seen as non-equilibrium analogues to the concepts of sequential and quasi-perfect equilibrium. Formally, this means that sequential equilibrium is obtained from sequential rationalizability by adding the requirement that there be mutual certain belief of the type profile, and likewise for quasi-perfect equilibrium. Similarly, rationalizability, permissibility, weak sequential rationalizability, and proper rationalizability may be viewed as non-equilibrium analogues to Nash equilibrium, strategic form perfect equilibrium, weak sequential equilibrium, and proper equilibrium, respectively. Tables 1 and 2 summarize the relations between the above mentioned equilibrium and rationalizability concepts, respectively, and provide relevant references.

We now proceed by showing the relations as stated in the second table. For this, it is useful to state the following definition: say that the mixed strategy $p_j^{t_i|t_j}$ is induced for t_j by t_i if $t_j \in T_j^{t_i}$ and, for all $s_j \in S_j$,

$$p_j^{t_i|t_j}(s_j) := \frac{\mu_\ell^{t_i}(s_j, t_j)}{\mu_\ell^{t_i}(S_j, t_j)},$$

where ℓ is the first level ℓ of λ^{t_i} for which $\mu_\ell^{t_i}(S_j, t_j) > 0$.

Table 1
Relationship between different equilibrium concepts

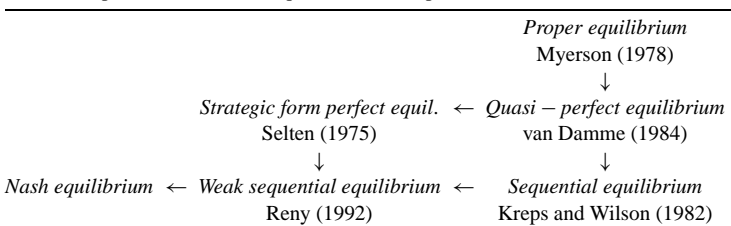


Table 2
Relationship between different rationalizability concepts

Common cert. belief that each player...	... believes the oppon. chooses rationally only in the whole game	... believes the oppon. chooses rationally at all reachable info. sets	... believes the oppon. chooses rationally at all info. sets
... is cautious and respects preferences	[n.a.]	[n.a.]	<i>Proper rationalizability</i> Schuhmacher (1999)
... is cautious	[n.a.]	<i>Permissibility</i> Börgers (1994) Brandenburger (1992) Dekel and Fudenberg (1990)	<i>Quasi-perfect rationalizability</i> (defined here)
... is not necessarily cautious	<i>Rationalizability</i> Bernheim (1984) ← Pearce (1984) Does not imply backward ind.	<i>Weak sequential rationalizability</i> Ben-Porath (1997) Does not imply backward ind.	<i>Sequential rationalizability</i> Dekel et al. (1999, 2002) Implies backward ind.

4.1. Properness

Say that each player respects opponent preferences in the sense of deeming one opponent strategy infinitely more likely than another if the opponent prefers the one to the other (cf. Blume et al., 1991b), as captured by the following event:

$$[resp_i] := \{(t_1, t_2) \in T_1 \times T_2 \mid \forall t'_j \in T_j^{t_i}, (s_j, t'_j) \gg^{t_i} (s'_j, t'_j) \text{ whenever } s_j \succ^{t'_j} s'_j\}.$$

Here, \gg^{t_i} is the “infinitely more likely” relation given t_i ’s system of conditional beliefs (cf. Appendix A), and, for each $t'_j \in T_j^{t_i}$, $\succ^{t'_j}$ denotes the complete and transitive preferences on S_j determined by j ’s vNM utility function v_j and t'_j ’s belief on $S_i \times T_i$. Write $[resp] := [resp_1] \cap [resp_2]$.

Building on Blume et al. (1991b, Proposition 5), Asheim (2001, Proposition 1) shows that proper equilibrium in two-player games can be characterized as a profile of induced mixed strategies at a type profile in $[resp] \cap [cau]$ where there is mutual certain belief of the type profile. Moreover, Asheim (2001, Proposition 3) can be used to show that Schuhmacher’s (1999) concept of properly rationalizable strategies corresponds to induced mixed strategies under common certain belief of $[resp] \cap [cau]$.

Any proper equilibrium in the strategic form corresponds to a quasi-perfect equilibrium in the extensive form (cf. van Damme, 1984). The following result (the proof of which is available on request) shows, by Proposition 7 and Asheim (2001, Proposition 1) this relationship between the equilibrium concepts and establishes, by Definition 8 and Asheim (2001, Proposition 3), the corresponding relationship between the rationalizability concepts. Furthermore, it means that the two cells in Table 2 to the left of ‘proper rationalizability’ are not applicable.

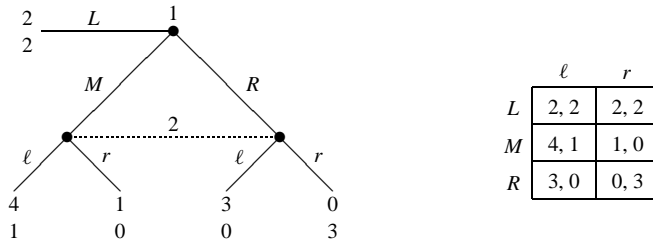


Fig. 2. Γ_2 and its strategic form.

Proposition 9. For any epistemic model and for each player i ,

$$[resp_i] \cap K_i[cau_j] \subseteq [isr_i].$$

From the proof of Mailath et al. (1997, Proposition 1) one can conjecture that quasi-perfect rationalizability in every extensive form corresponding to a given strategic game coincides with proper rationalizability in that game. However, for any given extensive form the set of proper rationalizable strategies can be a strict subset of the set of quasi-perfect rationalizable strategies, as illustrated by Γ_2 of Fig. 2.

Here, quasi-perfect rationalizability only precludes the play of R with positive probability. However, since M strongly dominates R , it follows that 2 prefers ℓ to r if she respects 1’s preferences. Hence, only ℓ with probability one is properly rationalizable for 2, which implies that only M with probability one is properly rationalizable for 1.

4.2. Weak sequentiality and permissibility

Say that a mixed strategy p_i is *weak sequentially rational* for t_i if,

$$\forall h \in H_i \quad \text{s.t.} \quad \text{supp } p_i \cap S_i(h) \neq \emptyset, \quad \text{supp } p_i \cap S_i(h) \subseteq C_i^{t_i}(h),$$

and define the event that player i is of a type that induces a weak sequentially rational mixed strategy for any opponent type that is deemed subjectively possible:

$$[iwr_i] := \{(t_1, t_2) \in T_1 \times T_2 \mid \forall t'_j \in T_j^{t_i}, p_j^{t_i|t'_j} \text{ is weak sequentially rational for } t'_j\}.$$

Write $[iwr] := [iwr_1] \cap [iwr_2]$.

Note that the mixed strategy induced for t'_j by t_i may be interpreted as t_i ’s *initial* belief about the behavior of t'_j . In contrast to the *behavior* strategy induced for t'_j by t_i , as defined in Section 3.1, the induced *mixed* strategy gives no information about how t_i revises his belief about the behavior of t'_j at player j information sets that are *unreachable* given t_i ’s initial belief about t'_j ’s behavior. Hence, if the true type profile (t_1, t_2) is in $[iwr_i]$, then player i believes that each subjectively possible opponent type t'_j chooses rationally at player j information sets that do not contradict t_i ’s initial belief about the behavior of t'_j . However, and this is the crucial difference when compared to the case where $(t_1, t_2) \in [isr_i]$: $(t_1, t_2) \in [iwr_i]$ entails *no* restriction on how t_i revises his beliefs about t'_j ’s behavior conditional on t'_j reaching “surprising” information sets. The above observation

explains why weak sequentially rationalizable (coined ‘weak extensive form rationalizable’ by Battigalli and Bonanno, 1999) strategies can be shown to correspond to induced mixed strategies under common certain belief of $[iwr]$.

Say that a mixed strategy p_i is *rational* for t_i if $p_i \in \Delta(C_i^{t_i})$, where

$$C_i^{t_i} := \{s_i \in S_i \mid s_i \succ^{t_i} s'_i \text{ for all } s'_i \in S_i\},$$

where \succ^{t_i} denotes the preferences on S_i determined by i ’s vNM utility function v_i and t_i ’s belief on $S_j \times T_j$, and define the event that player i is of a type that induces a rational mixed strategy for any opponent type that is deemed subjectively possible:

$$[ir_i] := \{(t_1, t_2) \in T_1 \times T_2 \mid \forall t'_j \in T_j, p_j^{t_i|t'_j} \text{ is rational for } t'_j\}.$$

Write $[ir] := [ir_1] \cap [ir_2]$. Then (see, e.g., Asheim and Dufwenberg, 2003, Proposition 5.2) permissible strategies correspond to induced mixed strategies under common certain belief of $[ir] \cap [cau]$. Of course, by instead considering common certain belief of $[ir]$, we obtain a characterization of ordinary rationalizability.

The following result (the proof of which is available on request) establishes the remaining relationships between the rationalizability concepts of Table 2.

Proposition 10. *For any epistemic model and for each player i ,*

$$[isr_i] \subseteq [iwr_i] \subseteq [ir_i] \quad \text{and} \quad [ir_i] \cap K_i[cau_j] \subseteq [iwr_i].$$

Since $[ir_i] \cap K_i[cau_j] \subseteq [iwr_i]$, the cell in Table 2 to the left of ‘permissibility’ is not applicable, and permissibility refines weak sequential rationalizability. Fig. 1 (cf. Section 1) shows that the inclusion can be strict: Permissibility, but not weak sequential rationalizability, precludes that player 1 plays U in Γ_1 . Since $[isr_i] \subseteq [iwr_i]$, Definition 5 entails that sequential rationalizability refines weak sequential rationalizability. Since $[isr_i] \subseteq [ir_i]$, Definition 8 entails that quasi-perfect rationalizability refines permissibility. That the two latter inclusions can be strict, is illustrated by Γ_3 of Fig. 3 (introduced by Reny, 1992, Fig. 1).

Here permissibility only precludes the play of D at 1’s second decision node. This can be established by applying the Dekel–Fudenberg (1990) procedure (i.e., one round of weak elimination followed by iterated strong elimination) which eliminates a strategy if and only if it is not permissible. Since all terminal nodes yield different payoffs, weak sequential rationalizability leads to the same conclusion.⁶ However, only the play

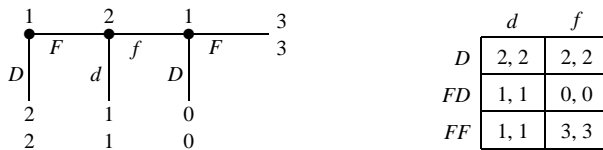


Fig. 3. Γ_3 and its strategic form.

⁶ To see how our characterization of weak sequential rationalizability is consistent with (D, d) in Γ_3 , let $T_1 = \{t_1\}$ with $\lambda^{t_1} = ((1, 0), (0, 1))$ (assigning probabilities to (d, t_2) and (f, t_2) respectively), and $T_2 = \{t_2\}$

of F with probability one at both of 1's decision nodes and the play of f at 2's single decision node are quasi-perfectly/sequentially rationalizable. This follows from Proposition 11 of Section 5, showing that the latter concepts imply the *backward induction procedure*.

4.3. Extensive form rationalizability

Extensive form rationalizability (EFR) (Pearce, 1984; Battigalli, 1997; Battigalli and Siniscalchi, 2002) is an iterative deletion procedure where, at any information set reached by a remaining strategy, any deleted strategy is deemed infinitely less likely than some remaining strategy. Even though EFR only requires players to choose rationally at reachable information sets and preference for cautious behavior is not imposed, EFR is different from weak sequential rationalizability. Unlike all concepts in Table 2, EFR yields *forward induction* in common examples like the 'Battle-of-the-Sexes-with-Outside-Option' and 'Burning Money' games.⁷ EFR also leads to the *backward induction outcome*. However, unlike proper, quasi-perfect and sequential rationalizability, EFR need not promote the backward induction procedure.

5. Relation to backward induction

The following result shows how sequential (and thus quasi-perfect and, by Proposition 9, proper) rationalizability implies the backward induction procedure in perfect information games. A finite extensive game Γ is of *perfect information* if, at any information set $h \in H_1 \cup H_2$, $h = \{x\}$; i.e., h contains only one node. It is *generic* if, for each i , $v_i(z) \neq v_i(z')$ whenever z and z' are different outcomes. A generic extensive game of perfect information has a unique subgame-perfect equilibrium in pure strategies. Moreover, in such games the backward induction procedure yields in any subgame the unique subgame-perfect equilibrium outcome.

Proposition 11. *Consider a finite generic extensive two-player game of perfect information Γ . If there exists an epistemic model with $(t_1, t_2) \in CK[isr]$ and, for each i , σ_i is induced for t_i by t_j , then $\sigma = (\sigma_1, \sigma_2)$ is the subgame-perfect equilibrium.*

Since sequentially rationalizable strategies always exist, there is an epistemic model with $(t_1, t_2) \in CK[isr]$, implying that the result of Proposition 11 is not empty.

with $\lambda^{t_2} = ((1, 0, 0), (0, 1, 0), (0, 0, 1))$ (assigning probabilities to (D, t_1) , (FD, t_1) , and (FF, t_1) , respectively). Then, independently of how ℓ^1 and ℓ^2 are specified, $(t_1, t_2) \in CK[iwr]$, and, for each i , p_i is induced for t_i by t_j , where $p_1(D) = 1$ and $p_2(d) = 1$.

⁷ By strengthening permissibility, Asheim and Dufwenberg (2003) define a rationalizability concept, *fully permissible sets*, which is different from those of Table 2 as well as EFR, as it yields forward induction, but does not always promote backward induction.

6. Concluding remarks

Throughout this paper, we have analyzed assumptions about players' *beliefs*, leading to events that are subsets of $T_1 \times T_2$. We can still make probabilistic statements about what a player "will do," by considering the beliefs of the other player.

For the concepts in the left and center columns of Table 2, we can do more than this, if we so wish. E.g., when characterizing weak sequential rationalizability, we can consider the event of rational pure choice at all reachable information sets, and assume that this event is commonly believed (where the term 'belief' is used in the sense of 'belief with probability one'). These assumptions yield subsets of $S_1 \times T_1 \times S_2 \times T_2$, leading to direct behavioral implications within the model.

This does not carry over to the concepts in the right column. It is problematic to define the event of rational pure choice at *all* information sets, since reaching a non-reachable information set may contradict rational choice at earlier information sets. Also, if we consider the event of (any kind of) rational pure choice, then we cannot use common *certain* belief, since this—combined with rational choice—would prevent well-defined conditional beliefs after irrational opponent choices. However, common belief (with probability one) of the event that each player believes his opponent chooses rationally at all information sets does *not* yield backward induction in generic perfect information games, as shown in the counterexample of Asheim (2002, Fig. 2). Common *certain* belief is essential for our analysis of the concepts in the right column of Table 2; this complicates obtaining direct behavioral implications.

In this paper, we have restricted our attention to games with two players. A natural question which arises is whether, and if so, how, the present analysis can be extended to the case of three or more players. In order to illustrate the potential difficulties of such an extension, consider a three player game in which player 3 has an information set h with two nodes, x and y , where x is preceded by the player 1 action a and the player 2 action c , and y is preceded by the player 1 action b and the player 2 action d . Suppose that player 3 views b and c as suboptimal choices, and hence player 3 deems a infinitely more likely than b , and deems d infinitely more likely than c . Then, player 3's LPS at h over player 1's strategy choice and player 3's LPS at h over player 2's strategy choice do not provide sufficient information to derive player 3's relative likelihoods attached to nodes x and y , and these relative likelihoods are crucial to assess player 3's rational behavior at h . Hence, in addition to the two LPSs mentioned above, we need another, aggregated, LPS for player 3 at h over his opponents' collective strategy profiles.

The key problem would then be what restrictions to impose upon the connection between the LPSs over individual strategies on the one hand and the aggregate LPS over strategy profiles on the other hand. Both classes of LPSs are needed, since the former are crucial in order to evaluate the beliefs about rationality of individual players, and the latter are needed in order to determine the conditional preferences of each player, as shown above. This issue is closely related to the problem of how to characterize consistency of assessments in algebraic terms, without the use of sequences (McLennan, 1989a; McLennan, 1989b; Battigalli, 1996; Kohlberg and Reny, 1997; Perea et al., 1997). In these papers, the consistency requirement for assessments has been characterized by means of conditional

probability systems, relative probability systems and lexicographic probability systems,⁸ satisfying some appropriate additional conditions. Perea et al. (1997), for instance, use a refinement of LPS in which, at every information set, not only an LPS over the available actions is defined, but moreover the relative likelihood level between actions is “quantified” by an additional parameter, whenever one action is deemed infinitely more likely than the other. This additional parameter makes it possible to derive a unique aggregate LPS over action profiles (and hence also over strategy profiles). A similar approach can be found in Govindan and Klumpp (2002). Such an approach could possibly be useful when extending the analysis in this paper to the case of more than two players. For the moment, we leave this issue for future research.

Acknowledgments

We have benefited from the detailed suggestions of a referee and the advice of an associate editor, extensive discussions with Drew Fudenberg and Joseph Greenberg, as well as comments from Pierpaolo Battigalli, James Bergin, Adam Brandenburger, Eddie Dekel, Takako Fujiwara-Greve, Andrew McLennan, and Larry Samuelson. Asheim gratefully acknowledges the hospitality of Harvard and Stanford Universities and financial support from the Research Council of Norway.

Appendix A. A representation result for SCLP

Consider the setting of Section 2.1. Write \mathbf{x}_E and \mathbf{y}_E for acts on $E \in 2^F \setminus \{\emptyset\}$. A binary relation on the set of acts on E is denoted by \succsim_E , where $\mathbf{x}_E \succsim_E \mathbf{y}_E$ means that \mathbf{x}_E is preferred or indifferent to \mathbf{y}_E . As usual, let \succ_E (preferred to) and \sim_E (indifferent to) denote the asymmetric and symmetric parts of \succsim_E . Assume that \succsim_E satisfies

Axiom 1 (Order). \succsim_E is complete and transitive.

Axiom 2 (Objective independence). $\mathbf{x}'_E \succ_E$ (respectively \sim_E) \mathbf{x}''_E iff $\gamma \mathbf{x}'_E + (1 - \gamma) \mathbf{y}_E \succ_E$ (respectively \sim_E) $\gamma \mathbf{x}''_E + (1 - \gamma) \mathbf{y}_E$ whenever $0 < \gamma < 1$ and \mathbf{y}_E is arbitrary.

for any $E \subseteq 2^F \setminus \{\emptyset\}$, and the following axiom if and only if $E \in F^*$,

Axiom 3 (Nontriviality). There exist \mathbf{x}_E and \mathbf{y}_E such that $\mathbf{x}_E \succ_E \mathbf{y}_E$.

where the numbering of axioms follows Blume et al. (1991a), henceforth referred to as BBDa. Say that $e \in F^*$ is deemed *infinitely more likely* than $f \in F$ ($e \gg f$) if

$$\mathbf{x}_{\{e,f\}} \succ_{\{e,f\}} \mathbf{y}_{\{e,f\}} \quad \text{whenever } \mathbf{x}_{\{e\}} \succ_{\{e\}} \mathbf{y}_{\{e\}}.$$

Let, for any $D \in 2^E \setminus \{\emptyset\}$, \mathbf{x}_D denote the restriction of \mathbf{x}_E to D . Define the conditional binary relation $\succsim_{E|D}$ by $\mathbf{x}'_E \succsim_{E|D} \mathbf{x}''_E$ if, for some \mathbf{y}_E , $(\mathbf{x}'_D, \mathbf{y}_{E \setminus D}) \succsim_E (\mathbf{x}''_D, \mathbf{y}_{E \setminus D})$. By Axioms 1 and 2, this definition does not depend on \mathbf{y}_E . Assume that \succsim_E satisfies

⁸ Cf. Hammond (1994) and Halpern (2003) for analyses of the relationship between these notions.

Axiom 4' (Conditional Archimedean property). $\forall e \in E, \exists 0 < \gamma < \delta < 1$ such that $\delta \mathbf{x}'_E + (1 - \delta) \mathbf{x}''_E \succ_{E|\{e\}} \mathbf{y}_E \succ_{E|\{e\}} \gamma \mathbf{x}'_E + (1 - \gamma) \mathbf{x}''_E$ whenever $\mathbf{x}'_E \succ_{E|\{e\}} \mathbf{y}_E \succ_{E|\{e\}} \mathbf{x}''_E$.

for any $E \subseteq 2^F \setminus \{\emptyset\}$.

The collection $\{\succ_E \mid E \in \mathcal{F}^*\}$ is called a system of conditional preferences on the set of states F . Assume that $\{\succ_E \mid E \in \mathcal{F}^*\}$ satisfies the following axioms:

Axiom 5* (Non-null state independence). $\mathbf{x}_{\{e\}} \succ_{\{e\}} \mathbf{y}_{\{e\}}$ iff $\mathbf{x}_{\{f\}} \succ_{\{f\}} \mathbf{y}_{\{f\}}$ whenever $e, f \in F^*$, and $\mathbf{x}_{\{e,f\}}$ and $\mathbf{y}_{\{e,f\}}$ satisfy $\mathbf{x}_{\{e,f\}}(e) = \mathbf{x}_{\{e,f\}}(f)$ and $\mathbf{y}_{\{e,f\}}(e) = \mathbf{y}_{\{e,f\}}(f)$.

Axiom 6* (Dynamic consistency). $\mathbf{x}_D \succ_D \mathbf{y}_D$ whenever $\mathbf{x}_E \succ_{E|D} \mathbf{y}_E$ and $\emptyset \neq D \subseteq E$.

Axiom 7* (Compatibility). There exists a binary relation \succ_F^* satisfying Axioms 1, 2, and 4' such that $\mathbf{x}_F \succ_{F|E}^* \mathbf{y}_F$ whenever $\mathbf{x}_E \succ_E \mathbf{y}_E$ and $\emptyset \neq E \subseteq F$.

Note that, for any event $E \in \mathcal{F}^*$, the decision-maker's actual conditional preferences over acts on E are given by \succ_E , while, e.g., $\succ_{F|E}$ and $\succ_{F|E}^*$ are auxiliary binary relations.

In their Theorem 3.1 BDDa show that a binary relation satisfying a set of axioms can be represented by a vNM utility function and an LPS. They impose Axioms 1–3 and 4' above, as well as Axiom 5 (non-null state independence), which coincides with our Axiom 5* in a setting where $\mathbf{x}_D \succ_D$ (respectively \sim_D) \mathbf{y}_D iff $\mathbf{x}_E \succ_{E|D}$ (respectively $\sim_{E|D}$) \mathbf{y}_E whenever $\emptyset \neq D \subseteq E$.

The following representation result extends BDDa's Theorem 3.1. For two utility vectors v and w , let $v \geq_L w$ denote that, whenever $w_\ell > v_\ell$, there exists $\ell' < \ell$ such that $v_{\ell'} > w_{\ell'}$, and let $>_L$ and $=_L$ denote the asymmetric and symmetric parts, respectively.

Proposition A.1. *The following two statements are equivalent:*

- (1) (a) \succ_E satisfies Axioms 1, 2, and 4' if $E \in 2^F \setminus \{\emptyset\}$, and Axiom 3 if and only if $E \in \mathcal{F}^*$, and
 (b) the system of conditional preferences $\{\succ_E \mid E \in \mathcal{F}^*\}$ satisfies Axioms 5*, 6*, and 7*.
- (2) There exist a vNM utility function $v: \Delta(Z) \rightarrow \mathbb{R}$ and an SCLP (λ, ℓ) on F with support F^* that satisfies, for any $E \in \mathcal{F}^*$,

$$\mathbf{x}_E \succ_E \mathbf{y}_E \quad \text{iff} \quad \left(\sum_{e \in E} \mu'_\ell(e) v(\mathbf{x}_E(e)) \right)_{\ell=1}^{\ell(E)|E} \geq_L \left(\sum_{e \in E} \mu'_\ell(e) v(\mathbf{y}_E(e)) \right)_{\ell=1}^{\ell(E)|E},$$

where $\lambda_{\ell(E)|E} = (\mu'_1, \dots, \mu'_{\ell(E)|E})$ is the conditional of $\lambda_{\ell(E)}$ on E .

Proof. (1) implies (2). Since \succ_E is trivial if $E \notin \mathcal{F}^*$, we may w.l.o.g. assume that Axiom 7* is satisfied with $\succ_{F|E}^*$ being trivial for any $E \notin \mathcal{F}^*$.

Consider any $e \in F^*$. Since $\succ_{\{e\}}$ satisfies Axioms 1–3, and 4' (implying BDDa's Axiom 4 since $\{e\}$ has only one state), it follows from von Neumann–Morgenstern expected

utility theory that there exists a vNM utility function $v_{\{e\}} : \Delta(Z) \rightarrow \mathbb{R}$ such that $v_{\{e\}}$ represents $\succsim_{\{e\}}$. By Axiom 5*, we may choose a common vNM utility function v to represent $\succsim_{\{e\}}$ for all $e \in F^*$. Since Axiom 7* implies, for any $e \in F^*$, $\succsim_{F|\{e\}}^*$ satisfies Axioms 1–3, and 4', and furthermore, $\mathbf{x}_F \succ_{F|\{e\}}^* \mathbf{y}_F$ whenever $\mathbf{x}_{\{e\}} \succ_{\{e\}} \mathbf{y}_{\{e\}}$, we obtain that v represents $\succsim_{F|\{e\}}^*$ for all $e \in F^*$. It now follows that \succsim_F^* satisfies Axiom 5 of BBDA.

By BBDA, Theorem 3.1, \succsim_F^* is represented by v and an LPS $\lambda = (\mu_1, \dots, \mu_L) \in \mathbf{L}\Delta(F)$ satisfying $\text{supp } \lambda = F^*$. Consider any $E \in \mathcal{F}^*$. If $\mathbf{x}_E \succ_E \mathbf{y}_E$ iff $\mathbf{x}_F \succ_{F|E}^* \mathbf{y}_F$, then

$$\mathbf{x}_E \succsim_E \mathbf{y}_E \quad \text{iff} \quad \left(\sum_{e \in E} \mu'_\ell(e) v(\mathbf{x}_E(e)) \right)_{\ell=1}^{L|E} \geq_L \left(\sum_{e \in E} \mu'_\ell(e) v(\mathbf{y}_E(e)) \right)_{\ell=1}^{L|E},$$

where $\lambda|_E = (\mu'_1, \dots, \mu'_{L|E})$ is the conditional of λ on E , implying that we can set $\ell(E) = L$.

Otherwise, let $\ell(E) \in \{0, \dots, L - 1\}$ be the maximum ℓ for which it holds that

$$\mathbf{x}_E \succ_E \mathbf{y}_E \quad \text{if} \quad \left(\sum_{e \in E} \mu'_{\ell'}(e) v(\mathbf{x}_E(e)) \right)_{\ell'=1}^{\ell|E} >_L \left(\sum_{e \in E} \mu'_{\ell'}(e) v(\mathbf{y}_E(e)) \right)_{\ell'=1}^{\ell|E},$$

where the r.h.s. is never satisfied if $\ell < \min\{\ell' | \text{supp } \lambda_{\ell'} \cap E \neq \emptyset\}$, entailing that the implication holds for any such ℓ . Define a set of pairs of acts on E, \mathcal{I} , as follows:

$$(\mathbf{x}_E, \mathbf{y}_E) \in \mathcal{I} \quad \text{iff} \quad \left(\sum_{e \in E} \mu'_\ell(e) v(\mathbf{x}_E(e)) \right)_{\ell=1}^{\ell(E)|E} =_L \left(\sum_{e \in E} \mu'_\ell(e) v(\mathbf{y}_E(e)) \right)_{\ell=1}^{\ell(E)|E},$$

with $(\mathbf{x}_E, \mathbf{y}_E) \in \mathcal{I}$ for any acts \mathbf{x}_E and \mathbf{y}_E on E if $\ell(E) < \min\{\ell | \text{supp } \lambda_\ell \cap E \neq \emptyset\}$. Note that \mathcal{I} is a convex set. To show that v and $\lambda_{\ell(E)|E}$ represent \succsim_E , we must establish that $\mathbf{x}_E \sim_E \mathbf{y}_E$ whenever $(\mathbf{x}_E, \mathbf{y}_E) \in \mathcal{I}$. Hence, suppose there exists $(\mathbf{x}_E, \mathbf{y}_E) \in \mathcal{I}$ such that $\mathbf{x}_E \succ_E \mathbf{y}_E$. It follows from the definition of $\ell(E)$ that there exists $(\mathbf{x}'_E, \mathbf{y}'_E) \in \mathcal{I}$ such that

$$\mathbf{x}'_E \succsim_E \mathbf{y}'_E \quad \text{and} \quad \sum_{e \in E} \mu_{\ell(E)+1}(e) v(\mathbf{x}'_E(e)) < \sum_{e \in E} \mu_{\ell(E)+1}(e) v(\mathbf{y}'_E(e)).$$

Objective independence of \succsim_E now implies that, if $0 < \gamma < 1$, then

$$\gamma \mathbf{x}_E + (1 - \gamma) \mathbf{x}'_E \succ_E \gamma \mathbf{y}_E + (1 - \gamma) \mathbf{x}'_E \succsim_E \gamma \mathbf{y}_E + (1 - \gamma) \mathbf{y}'_E;$$

hence, by transitivity of \succsim_E ,

$$\gamma \mathbf{x}_E + (1 - \gamma) \mathbf{x}'_E \succ_E \gamma \mathbf{y}_E + (1 - \gamma) \mathbf{y}'_E. \tag{A.1}$$

However, by choosing γ sufficiently small, we have that

$$\begin{aligned} & \sum_{e \in E} \mu_{\ell(E)+1}(e) v(\gamma \mathbf{x}_E(e) + (1 - \gamma) \mathbf{x}'_E(e)) \\ & < \sum_{e \in E} \mu_{\ell(E)+1}(e) v(\gamma \mathbf{y}_E(e) + (1 - \gamma) \mathbf{y}'_E(e)). \end{aligned}$$

Since \mathcal{I} is convex so that $(\gamma \mathbf{x}_E + (1 - \gamma) \mathbf{x}'_E, \gamma \mathbf{y}_E + (1 - \gamma) \mathbf{y}'_E) \in \mathcal{I}$, this implies that

$$\gamma \mathbf{x}_F + (1 - \gamma) \mathbf{x}'_F \prec_{F|E}^* \gamma \mathbf{y}_F + (1 - \gamma) \mathbf{y}'_F. \tag{A.2}$$

Since (A.1) and (A.2) contradict Axiom 7*, this shows that $\mathbf{x}_E \sim_E \mathbf{y}_E$ whenever $(\mathbf{x}_E, \mathbf{y}_E) \in \mathcal{I}$. This implies in turn that $\ell(E) \geq \min\{\ell \mid \text{supp } \lambda_\ell \cap E \neq \emptyset\}$ since \succsim_E is non-trivial. By Axiom 6*, $\ell(D) \geq \ell(E)$ whenever $\emptyset \neq D \subseteq E$. Finally, since, for any $e \in F^*$, $\mathbf{x}_{\{e\}} \succ_{\{e\}} \mathbf{y}_{\{e\}}$ iff $\mathbf{x}_F \succ_{F \setminus \{e\}}^* \mathbf{y}_F$, we have that $\ell(\{e\}) = L$, implying $\ell(\{e\}) \geq \ell$ whenever $e \in \text{supp } \mu_\ell$.

(2) implies (1). This follows from routine arguments. \square

Appendix B. Proofs of Propositions 4, 7, and 11

For the proofs of Propositions 4 and 7 we use two results from Blume et al. (1991b, henceforth referred to as BBDb). To state these results, we introduce the following notation. Let $\lambda = (\mu_1, \dots, \mu_L)$ be an LPS on a finite set F and let $r = (r_1, \dots, r_{L-1}) \in (0, 1)^{L-1}$. Then, $r \square \lambda$ denotes the probability distribution on F given by the nested convex combination

$$(1 - r_1)\mu_1 + r_1[(1 - r_2)\mu_2 + r_2[(1 - r_3)\mu_3 + r_3[\dots]]].$$

Lemma B.1. *Let $(p^n)_{n \in \mathbb{N}}$ be a sequence of probability distributions on a finite set F . Then, there exists a subsequence p^m of $(p^n)_{n \in \mathbb{N}}$, an LPS $\lambda = (\mu_1, \dots, \mu_L)$ and a sequence r^m of vectors in $(0, 1)^{L-1}$ converging to zero such that $p^m = r^m \square \lambda$ for all m .*

The following lemma is a variant of Proposition 1 in BBDb.

Lemma B.2. *Let player i 's preferences, \succsim^i , over acts on S_j be represented by v_i and $\lambda^i = (\mu_1^i, \dots, \mu_L^i) \in \mathbf{L}\Delta(S_j)$. Then,*

- (a) $s_i \succ^i s'_i$ if and only if for every sequence $(r^n)_{n \in \mathbb{N}}$ in $(0, 1)^{L-1}$ converging to zero there is a subsequence r^m such that

$$\sum_{s_j} (r^m \square \lambda^i)(s_j) u_i(s_i, s_j) > \sum_{s_j} (r^m \square \lambda^i)(s_j) u_i(s'_i, s_j)$$

for all m , and

- (b) the same result would hold if the phrase “for every sequence...” is replaced by “for some sequence...”.

Proof. (a) Suppose that $s_i \succ^i s'_i$. Then, there is some $k \in \{1, \dots, L\}$ such that

$$\sum_{s_j} \mu_\ell^i(s_j) u_i(s_i, s_j) = \sum_{s_j} \mu_\ell^i(s_j) u_i(s'_i, s_j) \tag{B.1}$$

for all $\ell < k$ and

$$\sum_{s_j} \mu_k^i(s_j) u_i(s_i, s_j) > \sum_{s_j} \mu_k^i(s_j) u_i(s'_i, s_j). \tag{B.2}$$

Let $(r^n)_{n \in \mathbb{N}}$ be a sequence in $(0, 1)^{L-1}$ converging to zero. By (B.1) and (B.2),

$$\sum_{s_j} (r^n \square \lambda^i)(s_j) u_i(s_i, s_j) > \sum_{s_j} (r^n \square \lambda^i)(s_j) u_i(s'_i, s_j)$$

if n is large enough. The other direction follows directly from the proof of Proposition 1 in BBDb. The proof of part (b) follows from the proof of Proposition 1 in BBDb. \square

For the proofs of Propositions 4 and 7 we need the following definitions. Let the LPS $\lambda^i = (\mu_1^i, \dots, \mu_L^i) \in \mathbf{L}\Delta(S_j)$ have full support on S_j . Say that the behavior strategy σ_j is induced by λ^i if for all $h \in H_j$ and $a \in A(h)$,

$$\sigma_j(h)(a) := \frac{\mu_\ell^i(S_j(h, a))}{\mu_\ell^i(S_j(h))},$$

where $\ell = \min\{\ell' \mid \text{supp } \lambda_{\ell'}^i \cap S_j(h) \neq \emptyset\}$. Moreover, say that player i 's beliefs over past opponent actions β_i are induced by λ_i if for all $h \in H_i$ and $x \in h$,

$$\beta_i(h)(x) := \frac{\mu_\ell^i(S_j(x))}{\mu_\ell^i(S_j(h))},$$

where $\ell = \min\{\ell' \mid \text{supp } \lambda_{\ell'}^i \cap S_j(h) \neq \emptyset\}$.

Proof of Proposition 4. (Only if) Let (σ, β) be a sequential equilibrium. Then (σ, β) is consistent and hence there is a sequence $(\sigma^n)_{n \in \mathbb{N}}$ of completely mixed behavior strategy profiles converging to σ such that the sequence $(\beta^n)_{n \in \mathbb{N}}$ of induced belief systems converges to β . For each i and all n , let $p_i^n \in \Delta(S_i)$ be the mixed representation of σ_i^n . By Lemma B.1, the sequence $(p_j^n)_{n \in \mathbb{N}}$ of probability distributions on S_j contains a subsequence p_j^m such that we can find an LPS $\lambda^i = (\mu_1^i, \dots, \mu_L^i)$ with full support on S_j and a sequence of vectors $r^m \in (0, 1)^{L-1}$ converging to zero with

$$p_j^m = r^m \square \lambda^i$$

for all m . W.l.o.g., we assume that $p_j^n = r^n \square \lambda^i$ for all $n \in \mathbb{N}$.

We first show that λ^i induces the behavior strategy σ_j . Let $\tilde{\sigma}_j$ be the behavior strategy induced by λ^i . By definition, $\forall h \in H_j, \forall a \in A(h)$,

$$\begin{aligned} \tilde{\sigma}_j(h)(a) &= \frac{\mu_\ell^i(S_j(h, a))}{\mu_\ell^i(S_j(h))} = \lim_{n \rightarrow \infty} \frac{(r^n \square \lambda^i)(S_j(h, a))}{(r^n \square \lambda^i)(S_j(h))} \\ &= \lim_{n \rightarrow \infty} \frac{p_j^n(S_j(h, a))}{p_j^n(S_j(h))} = \lim_{n \rightarrow \infty} \sigma_j^n(h)(a) = \sigma_j(h)(a), \end{aligned}$$

where $\ell = \min\{\ell' \mid \text{supp } \lambda_{\ell'}^i \cap S_j(h) \neq \emptyset\}$. For the fourth equation we used the fact that p_j^n is the mixed representation of σ_j^n . Hence, for each i , λ^i induces σ_j .

We then show that λ^i induces the beliefs β_i . Let $\tilde{\beta}_i$ be player i 's beliefs over past opponent actions induced by λ^i . By definition, $\forall h \in H_i, \forall x \in h$,

$$\begin{aligned} \tilde{\beta}_i(h)(x) &= \frac{\mu_\ell^i(S_j(x))}{\mu_\ell^i(S_j(h))} = \lim_{n \rightarrow \infty} \frac{r^n \square \lambda^i(S_j(x))}{r^n \square \lambda^i(S_j(h))} \\ &= \lim_{n \rightarrow \infty} \frac{p_j^n(S_j(x))}{p_j^n(S_j(h))} = \lim_{n \rightarrow \infty} \beta_i^n(h)(x) = \beta_i(h)(x), \end{aligned}$$

where $\ell = \min\{\ell' \mid \text{supp } \lambda_{\ell'}^i \cap S_j(h) \neq \emptyset\}$. For the fourth equality we used the facts that p_j^n is the mixed representation of σ_j^n and β_i^n is induced by σ_j^n . Hence, for each i , λ^i induces β_i .

We now define the following epistemic model. Let $T_1 = \{t_1\}$ and $T_2 = \{t_2\}$. Let, for each i , (1) (λ^i, ℓ^i) be the SCLP with support $S_j \times \{t_j\}$ where λ^i coincides with the λ^i constructed above, and (2) $\ell^i(E_j) = \min\{\ell \mid \text{supp } \lambda_{\ell}^i \cap E_j \neq \emptyset\}$ for all $(\emptyset \neq) E_j \subseteq S_j \times \{t_j\}$. Then, it is clear that there is mutual certain belief of $\{(t_1, t_2)\}$ at (t_1, t_2) , and for each i , σ_i is induced for t_i by t_j . It remains to show that $(t_1, t_2) \in [isr]$.

For this, it is sufficient to show, for each i , that σ_i is sequentially rational for t_i . Suppose not. By the choice of ℓ^i , it then follows that there is some information set $h \in H_i$ and some mixed strategy $p_i \in \Delta(S_i(h))$ that is outcome-equivalent to $\sigma_i|_h$ such that there exist $s_i \in S_i(h)$ with $p_i(s_i) > 0$ and $s'_i \in S_i(h)$ having the property that

$$u_i(s_i, \mu_{\ell}^i|_{S_j(h)}) < u_i(s'_i, \mu_{\ell}^i|_{S_j(h)}),$$

where $\ell = \min\{\ell' \mid \text{supp } \lambda_{\ell'}^i \cap (S_j(h) \times \{t_j\}) \neq \emptyset\}$ and $\mu_{\ell}^i|_{S_j(h)} \in \Delta(S_j(h))$ is the conditional probability distribution on $S_j(h)$ induced by μ_{ℓ}^i . Recall that μ_{ℓ}^i is the ℓ -th level of the LPS λ^i . Since the beliefs β_i and the behavior strategy σ_j are induced by λ^i , it follows that $u_i(s_i, \mu_{\ell}^i|_{S_j(h)}) = u_i(s_i, \sigma_j; \beta_i)|_h$ and $u_i(s'_i, \mu_{\ell}^i|_{S_j(h)}) = u_i(s'_i, \sigma_j; \beta_i)|_h$ and hence

$$u_i(s_i, \sigma_j; \beta_i)|_h < u_i(s'_i, \sigma_j; \beta_i)|_h,$$

which is a contradiction to the fact that (σ, β) is sequentially rational.

(If) Suppose that there is an epistemic model with $(t_1, t_2) \in [isr]$ such that there is mutual certain belief of $\{(t_1, t_2)\}$ at (t_1, t_2) , and for each i , σ_i is induced for t_i by t_j . We show that $\sigma = (\sigma_1, \sigma_2)$ can be extended to a sequential equilibrium.

For each i , let $\lambda^i = (\mu_1^i, \dots, \mu_L^i) \in \mathbf{L}\Delta(S_j)$ be the LPS coinciding with λ^i , and let β_i be player i 's beliefs over past opponent choices induced by λ^i . Write $\beta = (\beta_1, \beta_2)$. We first show that (σ, β) is consistent.

Choose sequences $(r^n)_{n \in \mathbb{N}}$ in $(0, 1)^{L-1}$ converging to zero and let the sequences $(p_j^n)_{n \in \mathbb{N}}$ of mixed strategies be given by $p_j^n = r^n \square \lambda^i$ for all n . Since λ^i has full support on S_j for every n , p_j^n is completely mixed. For every n , let σ_j^n be a behavior representation of p_j^n and let β_i^n be the beliefs induced by σ_j^n . We show that $(\sigma_j^n)_{n \in \mathbb{N}}$ converges to σ_j and that $(\beta_i^n)_{n \in \mathbb{N}}$ converges to β_i , which imply consistency of (σ, β) .

Note that the inducement of σ_j by t_i depends on λ^i through, for each $h \in H_j, \mu_{\ell}^i$, where $\ell = \min\{\ell' \mid \text{supp } \lambda_{\ell'}^i \cap (S_j(h) \times \{t_j\}) \neq \emptyset\}$. This implies that σ_j is induced by λ^i . Since σ_j^n is a behavior representation of p_j^n and σ_j is induced by λ^i , we have, $\forall h \in H_j, \forall a \in A(h)$,

$$\begin{aligned} \lim_{n \rightarrow \infty} \sigma_j^n(h)(a) &= \lim_{n \rightarrow \infty} \frac{p_j^n(S_j(h, a))}{p_j^n(S_j(h))} = \lim_{n \rightarrow \infty} \frac{r^n \square \lambda^i(S_j(h, a))}{r^n \square \lambda^i(S_j(h))} \\ &= \frac{\mu_{\ell}^i(S_j(h, a))}{\mu_{\ell}^i(S_j(h))} = \sigma_j(h)(a), \end{aligned}$$

where $\ell = \min\{\ell' \mid \text{supp } \lambda_{\ell'}^i \cap S_j(h) \neq \emptyset\}$. Hence, $(\sigma_j^n)_{n \in \mathbb{N}}$ converges to σ_j .

Since β_i^n is induced by σ_j^n and σ_j^n is a behavior representation of p_j^n , and furthermore, β_i is induced by λ^i , we have, $\forall h \in H_i, \forall x \in h$,

$$\begin{aligned} \lim_{n \rightarrow \infty} \beta_i^n(h)(x) &= \lim_{n \rightarrow \infty} \frac{p_j^n(S_j(x))}{p_j^n(S_j(h))} = \lim_{n \rightarrow \infty} \frac{r^n \square \lambda^i(S_j(x))}{r^n \square \lambda^i(S_j(h))} \\ &= \frac{\mu_\ell^i(S_j(x))}{\mu_\ell^i(S_j(h))} = \beta_i(h)(x), \end{aligned}$$

where $\ell = \min\{\ell' \mid \text{supp} \lambda_{\ell'}^i \cap S_j(h) \neq \emptyset\}$. Hence, $(\beta_i^n)_{n \in \mathbb{N}}$ converges to β_i .

This establishes that (σ, β) is consistent.

It remains to show that for each i and $\forall h \in H_i$,

$$u_i(\sigma_i, \sigma_j; \beta_i)|_h = \max_{\sigma'_i} u_i(\sigma'_i, \sigma_j; \beta_i)|_h.$$

Suppose not. Then, $u_i(\sigma_i, \sigma_j; \beta_i)|_h < u_i(\sigma'_i, \sigma_j; \beta_i)|_h$ for some $h \in H_i$ and some σ'_i . Let $p_i \in \Delta(S_i(h))$ be outcome-equivalent to $\sigma_i|_h$. Then, there is some $s_i \in S_i(h)$ with $p_i(s_i) > 0$ and some $s'_i \in S_i(h)$ such that

$$u_i(s_i, \sigma_j; \beta_i)|_h < u_i(s'_i, \sigma_j; \beta_i)|_h.$$

Since the beliefs β_i and the behavior strategy σ_j are induced by λ^i , it follows (using the notation that has been introduced in the ‘only if’ part of this proof) that $u_i(s_i, \sigma_j; \beta_i)|_h = u_i(s_i, \mu_\ell^i|_{S_j(h)})$ and $u_i(s'_i, \sigma_j; \beta_i)|_h = u_i(s'_i, \mu_\ell^i|_{S_j(h)})|_h$ and hence

$$u_i(s_i, \mu_\ell^i|_{S_j(h)}) < u_i(s'_i, \mu_\ell^i|_{S_j(h)}),$$

which contradicts the fact that σ_i is sequentially rational for t_i . This completes the proof of this proposition. \square

Proof of Proposition 7. (Only if) Let (σ_1, σ_2) be a quasi-perfect equilibrium. By definition, there is a sequence $(\sigma^n)_{n \in \mathbb{N}}$ of completely mixed behavior strategy profiles converging to σ such that for each i and every $n \in \mathbb{N}$ and $h \in H_i$,

$$u_i(\sigma_i, \sigma_j^n)|_h = \max_{\sigma'_i} u_i(\sigma'_i, \sigma_j^n)|_h.$$

For each j and every n , let p_j^n be the mixed representation of σ_j^n . By Lemma B.1, the sequence $(p_j^n)_{n \in \mathbb{N}}$ of probability distributions on S_j contains a subsequence p_j^m such that we can find an LPS $\lambda^i = (\mu_1^i, \dots, \mu_L^i)$ with full support on S_j and a sequence of vectors $r^m \in (0, 1)^{L-1}$ converging to zero with

$$p_j^m = r^m \square \lambda^i$$

for all m . W.l.o.g., we assume that $p_j^n = r^n \square \lambda^i$ for all $n \in \mathbb{N}$.

By the same argument as in the proof of Proposition 4, it follows that λ^i induces the behavior strategy σ_j . Now, we define an epistemic model as follows. Let $T_1 = \{t_1\}$ and $T_2 = \{t_2\}$. Let, for each i , (1) λ^i be the LPS on $S_j \times \{t_j\}$ which coincides with λ^i , and (2) $\ell^{t_i}(S_j \times \{t_j\}) = L$. Then, it is clear that there is mutual certain belief of $\{(t_1, t_2)\}$ at (t_1, t_2) , and for each i , σ_i is induced for t_i by t_j . It remains to show that $(t_1, t_2) \in [isr] \cap [cau]$.

Since, obviously, $(t_1, t_2) \in [cau]$, it suffices to show, for each i , that σ_i is sequentially rational for t_i . Fix a player i and let $h \in H_i$ be given. Let $p_i \in \Delta(S_i(h))$ be outcome-equivalent to $\sigma_i|_h$ and let p_j^n be the mixed representation of σ_j^n . Then, since (σ_1, σ_2) is a quasi-perfect equilibrium, it follows that

$$u_i(p_i, p_j^n|_h) = \max_{p'_i \in \Delta(S_i(h))} u_i(p'_i, p_j^n|_h)$$

for all n . Hence, $p_i(s_i) > 0$ implies that

$$\sum_{s_j \in S_j(h)} p_j^n|_h(s_j) u_i(s_i, s_j) = \max_{s'_i \in S_i(h)} \sum_{s_j \in S_j(h)} p_j^n|_h(s_j) u_i(s'_i, s_j) \tag{B.3}$$

for all n . Let $\lambda^i|_h$ be the conditional of λ^i on $S_j(h)$. Since $p_j^n = r^n \square \lambda^i$ for all n there exist vectors $r^n|_h$ converging to zero such that $p_j^n|_h = r^n|_h \square \lambda^i|_h$ for all n . Together with Eq. (B.3) we obtain that $p_i(s_i) > 0$ implies

$$\sum_{s_j \in S_j(h)} (r^n|_h \square \lambda^i|_h)(s_j) u_i(s_i, s_j) = \max_{s'_i \in S_i(h)} \sum_{s_j \in S_j(h)} (r^n|_h \square \lambda^i|_h)(s_j) u_i(s'_i, s_j). \tag{B.4}$$

We show that $p_i(s_i) > 0$ implies $s_i \in C_i^{t_i}(h)$. Suppose that $s_i \in S_i(h) \setminus C_i^{t_i}(h)$. Then, there is some $s'_i \in S_i(h)$ with $s'_i \succ_h^i s_i$. By applying Lemma B.2(a) in the case of acts on $S_j(h)$, it follows that $r^n|_h$ has a subsequence $r^m|_h$ for which

$$\sum_{s_j \in S_j(h)} (r^m|_h \square \lambda^i|_h)(s_j) u_i(s'_i, s_j) > \sum_{s_j \in S_j(h)} (r^m|_h \square \lambda^i|_h)(s_j) u_i(s_i, s_j)$$

for all m , which is a contradiction to (B.4). Hence, $s_i \in C_i^{t_i}(h)$ whenever $p_i(s_i) > 0$, which implies that $p_i \in \Delta(C_i^{t_i}(h))$. Hence, $\sigma_i|_h$ is outcome equivalent to some $p_i \in \Delta(C_i^{t_i}(h))$. This holds for every $h \in H_i$, and hence σ_i is sequentially rational for t_i .

(If) Suppose, there is an epistemic model with $(t_1, t_2) \in [isr] \cap [cau]$ such that there is mutual certain belief of $\{(t_1, t_2)\}$ at (t_1, t_2) , and for both i , σ_i is induced for t_i by t_j . We show that (σ_1, σ_2) is a quasi-perfect equilibrium.

For each i , let $\lambda^i = (\mu_1^i, \dots, \mu_L^i) \in \mathbf{L}\Delta(S_j)$ be the LPS coinciding with λ^{t_i} and let, for every $h \in H_i$, $\lambda^i|_h$ be the conditional of λ^i on $S_j(h)$. Since $(t_1, t_2) \in [cau_i]$, $\lambda^i|_h$ describes i 's conditional belief on $S_j(h)$. Choose sequences $(r^n)_{n \in \mathbb{N}}$ in $(0, 1)^{L-1}$ converging to zero and let the sequences $(p_j^n)_{n \in \mathbb{N}}$ of mixed strategies be given by $p_j^n = r^n \square \lambda^i$ for all n . Since λ^i has full support on S_j for every n , p_j^n is completely mixed. For every n , let σ_j^n be a behavior representation of p_j^n . Since λ^i induces σ_j , it follows that $(\sigma_j^n)_{n \in \mathbb{N}}$ converges to σ_j ; this is shown explicitly under the ‘if’ part of Proposition 4. Hence, to establish that (σ_1, σ_2) is a quasi-perfect equilibrium, we must show that, for each i and $\forall n \in \mathbb{N}$ and $\forall h \in H_i$,

$$u_i(\sigma_i, \sigma_j^n)|_h = \max_{\sigma'_i} u_i(\sigma'_i, \sigma_j^n)|_h. \tag{B.5}$$

Fix a player i and an information set $h \in H_i$. Let $p_i \in \Delta(S_i(h))$ be outcome-equivalent to $\sigma_i|_h$. Then, Eq. (B.5) is equivalent to

$$u_i(p_i, p_j^n|_h) = \max_{p'_i \in \Delta(S_i(h))} u_i(p'_i, p_j^n|_h)$$

for all n . Hence, we must show that $p_i(s_i) > 0$ implies that

$$\sum_{s_j \in S_j(h)} p_j^n |_h(s_j) u_i(s_i, s_j) = \max_{s'_i \in S_i(h)} \sum_{s_j \in S_j(h)} p_j^n |_h(s_j) u_i(s'_i, s_j) \tag{B.6}$$

for all n . In fact, it suffices to show this equation for infinitely many n , since in this case we can choose a subsequence for which the above equation holds, and this would be sufficient to show that (σ_1, σ_2) is a quasi-perfect equilibrium.

Since, by assumption, σ_i is sequentially rational for t_i , $\sigma_i |_h$ is outcome equivalent to some mixed strategy in $\Delta(C_i^{t_i}(h))$. Hence, $p_i \in \Delta(C_i^{t_i}(h))$. Let $p_i(s_i) > 0$. By construction, $s_i \in C_i^{t_i}(h)$. Suppose that s_i would not satisfy (B.6) for infinitely many n . Then, there exists some $s'_i \in S_i(h)$ such that

$$\sum_{s_j \in S_j(h)} p_j^n |_h(s_j) u_i(s_i, s_j) < \sum_{s_j \in S_j(h)} p_j^n |_h(s_j) u_i(s'_i, s_j)$$

for infinitely many n . Assume, w.l.o.g., that it is true for all n . Let $\lambda^i |_h$ be the conditional of λ^i on $S_j(h)$. Since $p_j^n = r^n \square \lambda^i$ for all n there exist vectors $r^n |_h$ converging to zero such that $p_j^n |_h = r^n |_h \square \lambda^i |_h$ for all n . This implies that

$$\sum_{s_j \in S_j(h)} (r^n |_h \square \lambda^i |_h)(s_j) u_i(s_i, s_j) < \sum_{s_j \in S_j(h)} (r^n |_h \square \lambda^i |_h)(s_j) u_i(s'_i, s_j)$$

for all n . By applying Lemma B.2(b) in the case of acts on $S_j(h)$, it follows that s'_i is strictly preferred by t_i to s_i at h , which is a contradiction to the fact that $s_i \in C_i^{t_i}(h)$. Hence, $p_i(s_i) > 0$ implies (B.6) for infinitely many n , and as a consequence, (σ_1, σ_2) is a quasi-perfect equilibrium. \square

Proof of Proposition 11. For this proof we must derive some properties of the certain belief operator (cf. Section 2.4). It is easy to check that $K_i(T) = T$ and $K_i\emptyset = \emptyset$, and, for any events A and B , $K_i A \cap K_i B = K_i(A \cap B)$, $K_i A \subseteq K_i K_i A$, and $\neg K_i A \subseteq K_i(\neg K_i A)$, implying that, for any event A , $K_i A = K_i K_i A$. Write $K^0 A := A$ and, for each $g \geq 1$, $K^g A := K K^{g-1} A$. Since $K_i(A \cap B) = K_i A \cap K_i B$ and $K_i K_i A = K_i A$, it follows $\forall g \geq 2$, $K^g A = K_1 K^{g-1} A \cap K_2 K^{g-1} A \subseteq K_1 K_1 K^{g-2} A \cap K_2 K_2 K^{g-2} A = K_1 K^{g-2} A \cap K_2 K^{g-2} A = K^{g-1} A$. Even though the truth axiom ($K_i A \subseteq A$) is not satisfied, the present paper considers certain belief only of events $A \subseteq T$ that can be written as $A = A_1 \cap A_2$ where, for each i , $A_i = \text{proj}_{T_i} A_i \times T_j$. Since each player certainly believes his own type, mutual certain belief of any such event A implies that A is true: $K A = K_1 A \cap K_2 A \subseteq K_1 A_1 \cap K_2 A_2 = A_1 \cap A_2 = A$ since, for each i , $K_i A_i = A_i$. Hence, (1) $\forall g \geq 1$, $K^g A \subseteq K^{g-1} A$, and (2) $\exists g' \geq 0$ such that $K^g A = K A \cap K K A \cap K K K A \cap \dots = C K A$ for $g \geq g'$ since T is finite.

In a perfect information game, the action $a \in A(h)$ taken at the information set h determines the immediate succeeding information set, which can thus be denoted (h, a) . Also, any information set $h \in H_1 \cup H_2$ determines a subgame. Set $H^{-1} = Z$ (i.e., the set of terminal nodes) and determine H^g for $g \geq 0$ by induction: $h \in H^g$ if and only if h satisfies

$$\max\{g' \mid \exists h' \in H^{g'} \text{ and } a \in A(h) \text{ such that } h' = (h, a)\} = g - 1.$$

In words, $h \in H^g$ if and only if g is the maximal number of decision nodes between h and a terminal node in the subgame determined by h . If σ is a profile of behavior strategies and $h \in H_1 \cup H_2$, denote by $\sigma|_h$ the strategy profile with the following properties:

- (1) at information sets preceding h , $\sigma|_h$ determines with probability one the unique action leading to h , and
- (2) at all other information sets, $\sigma|_h$ coincides with σ .

Say that σ' is *outcome-equivalent* to σ'' if σ' and σ'' induce the same probability distribution over terminal nodes.

In view of properties of the certain belief operator, it is sufficient to show for any $g = 0, \dots, \max\{g' \mid H^{g'} \neq \emptyset\}$ that if there exists an epistemic model with $(t_1, t_2) \in K^g[isr]$ and, for each i , σ_i is induced for t_i by t_j , then, $\forall h \in H^g$, $\sigma|_h$ is outcome-equivalent to $\sigma^*|_h$, where $\sigma^* = (\sigma_1^*, \sigma_2^*)$ denotes the subgame-perfect equilibrium. This is established by induction.

($g = 0$) Let $(t_1, t_2) \in K^0[isr] = [isr]$ and, for each i , σ_i be induced for t_i by t_j . Let $h \in H^0$ and assume w.l.o.g. that $h \in H_i$. Since $(t_1, t_2) \in [isr_j]$ and j takes no action at h , $\sigma|_h$ is outcome equivalent to $\sigma^*|_h$.

($g = 1, \dots, \max\{g' \mid H^{g'} \neq \emptyset\}$) Suppose that it has been established for $g' = 0, \dots, g - 1$ that if there exists an epistemic model with $(t_1, t_2) \in K^{g'}[isr]$ and, for each i , σ_i is induced for t_i by t_j , then, $\forall h' \in H^{g'}$, $\sigma|_{h'}$ is outcome-equivalent to $\sigma^*|_{h'}$. Let $(t_1, t_2) \in K^g[isr]$ and, for each i , σ_i be induced for t_i by t_j . Let $h \in H^g$ and assume w.l.o.g. that $h \in H_i$. Since $(t_1, t_2) \in K_i K^{g-1}[isr]$, it follows from the premise of the inductive step that t_i 's SCLP $(\lambda^{t_i}, \ell^{t_i})$ satisfies, $\forall t'_j \in T_j^{t_i}$, $\forall h' \in H_j$ succeeding h , and $\forall a' \in A(h')$,

$$\frac{\mu_\ell^{t_i}(S_j(h', a'), t'_j)}{\mu_\ell^{t_i}(S_j(h'), t'_j)} = \sigma_j^*(h')(a'),$$

where ℓ is the first level ℓ of λ^{t_i} for which $\mu_\ell^{t_i}(S_j(h'), t'_j) > 0$. Since Γ is generic, σ_i is sequentially rational for t_i only if $\sigma_i(h) = \sigma_i^*(h)$. Since $(t_1, t_2) \in [isr_j]$ and j takes no action at h , it follows from the premise that $\sigma|_h$ is outcome-equivalent to $\sigma^*|_h$. \square

References

- Anscombe, F.J., Aumann, R., 1963. A definition of subjective probability. *Annals Math. Stat.* 34, 199–205.
- Asheim, G.B., 2001. Proper rationalizability in lexicographic beliefs. *Int. J. Game Theory* 30, 453–478.
- Asheim, G.B., 2002. On the epistemic foundation for backward induction. *Math. Soc. Sci.* 44, 121–144.
- Asheim, G.B., Dufwenberg, M., 2003. Admissibility and common belief. *Games Econ. Behav.* 42, 208–234.
- Aumann, R.J., Brandenburger, A., 1995. Epistemic conditions for Nash equilibrium. *Econometrica* 63, 1161–1180.
- Battigalli, P., 1996. Strategic independence and perfect Bayesian equilibria. *J. Econ. Theory* 70, 201–234.
- Battigalli, P., 1997. On rationalizability in extensive games. *J. Econ. Theory* 74, 40–61.
- Battigalli, P., Bonanno, G., 1999. Recent results on belief, knowledge and the epistemic foundations of game theory. *Res. Econ.* 53, 149–225.
- Battigalli, P., Siniscalchi, M., 2002. Strong belief and forward induction reasoning. *J. Econ. Theory* 106, 356–391.
- Ben-Porath, E., 1997. Rationality, Nash equilibrium, and backwards induction in perfect information games. *Rev. Econ. Stud.* 64, 23–46.

- Bernheim, D., 1984. Rationalizable strategic behavior. *Econometrica* 52, 1007–1028.
- Blume, L., Brandenburger, A., Dekel, E., 1991a. Lexicographic probabilities and choice under uncertainty. *Econometrica* 59, 61–79.
- Blume, L., Brandenburger, A., Dekel, E., 1991b. Lexicographic probabilities and equilibrium refinements. *Econometrica* 59, 81–98.
- Börgers, T., 1994. Weak dominance and approximate common knowledge. *J. Econ. Theory* 64, 265–276.
- Brandenburger, A., 1992. Lexicographic probabilities and iterated admissibility. In: Dasgupta, P., Gale, D., Hart, O., Maskin, E. (Eds.), *Economic Analysis of Markets and Games*. MIT Press, Cambridge, MA, pp. 282–290.
- Dekel, E., Fudenberg, D., 1990. Rational behavior with payoff uncertainty. *J. Econ. Theory* 52, 243–267.
- Dekel, E., Fudenberg, D., Levine, D.K., 1999. Payoff information and self-confirming equilibrium. *J. Econ. Theory* 89, 165–185.
- Dekel, E., Fudenberg, D., Levine, D.K., 2002. Subjective uncertainty over behavior strategies: A correction. *J. Econ. Theory* 104, 473–478.
- Govindan, S., Klumpp, T., 2002. Perfect equilibrium and lexicographic beliefs. *Int. J. Game Theory* 31, 229–243.
- Greenberg, J., 1996. *Towering Over Babel: Worlds Apart but Acting Together*. McGill University.
- Greenberg, J., Gupta, S., Luo, X., 2003. *Towering Over Babel: Worlds Apart but Acting Together*. McGill University.
- Halpern, J.Y., 2003. *Lexicographic Probability, Conditional Probability, and Nonstandard Probability*. Cornell University.
- Hammond, P.J., 1994. Elementary non-archimedean representations of probability for decision theory and games. In: Humphreys, P. (Ed.), *Patrick Suppes: Scientific Philosopher*. In: *Probability and Probabilistic Causality*, vol. 1. Kluwer Academic, Dordrecht, pp. 25–59.
- Kohlberg, E., Reny, P.J., 1997. Independence on relative probability spaces and consistent assessments in game trees. *J. Econ. Theory* 75, 280–313.
- Kreps, D.M., Wilson, R., 1982. Sequential equilibria. *Econometrica* 50, 863–894.
- McLennan, A., 1989a. The space of conditional systems is a ball. *Int. J. Game Theory* 18, 125–139.
- McLennan, A., 1989b. Consistent conditional systems in noncooperative game theory. *Int. J. Game Theory* 18, 141–174.
- Mailath, G., Samuelson, L., Swinkels, J., 1997. How proper is sequential equilibrium? *Games Econ. Behav.* 18, 193–218.
- Myerson, R., 1978. Refinement of the Nash equilibrium concept. *Int. J. Game Theory* 7, 73–80.
- Myerson, R., 1986. Multistage games with communication. *Econometrica* 54, 323–358.
- Myerson, R., 1991. *Game Theory*. Harvard Univ. Press, Cambridge, MA.
- Pearce, D.G., 1984. Rationalizable strategic behavior and the problem of perfection. *Econometrica* 52, 1029–1050.
- Perea, A., 2002. Forward induction and the minimum revision principle. Meteor research memorandum 02/010, University of Maastricht.
- Perea, A., 2003. Rationalizability and minimal complexity in dynamic games. Meteor research memorandum 03/030, University of Maastricht.
- Perea, A., Jansen, M., Peters, H., 1997. Characterization of consistent assessments in extensive form games. *Games Econ. Behav.* 21, 238–252.
- Reny, P.J., 1992. Backward induction, normal form perfection and explicable equilibria. *Econometrica* 60, 627–649.
- Schuhmacher, F., 1999. Proper rationalizability and backward induction. *Int. J. Game Theory* 28, 599–615.
- Selten, R., 1975. Reexamination of the perfectness concept for equilibrium points in extensive games. *Int. J. Game Theory* 4, 25–55.
- van Damme, E., 1984. A relation between perfect equilibria in extensive form games and proper equilibria in normal form games. *Int. J. Game Theory* 13, 1–13.