Admissibility and Common Knowledge*

by

Geir B. Asheim** & Martin Dufwenberg***

This version: January 1996

ABSTRACT:

The implications of assuming that it is commonly known that players consider only admissible best responses are investigated. Within a states-of-the-world model where a state, for each player, determines a strategy set rather than a strategy the concept of *fully permissible sets* is defined. General existence is established, and a finite algorithm (eliminating strategy sets instead of strategies) is provided. The concept refines rationalizability as well as the Dekel-Fudenberg procedure, and captures a notion of forward induction. When players consider all best responses, the same framework can be used to define the concept of *rationalizable sets*, which characterizes rationalizability.

---

* This paper supersedes our earlier paper "Rational Reasoning and Rationalizable Sets". We thank Pierpaolo Battigalli, Elchanan Ben-Porath, Ken Binmore, Giacomo Bonanno, Eddie Dekel, Christian Ewerhart, Ebbe Groes, Faruk Gul, Hans Krogh Hvide, George Mailath, Roger Myerson, Phil Reny, Larry Samuelson, Jeroen Swinkels, and Jörgen Weibull for valuable discussions and comments, as well as a referee for constructive and very helpful suggestions. Asheim gratefully acknowledges the hospitality of Northwestern University and financial support from the Norwegian Research Council. Finally, the acknowledgments in Asheim (1994) and Dufwenberg (1994) apply also here; in particular, thanks go to Dilip Abreu, Yossi Greenberg, Michihiro Kandori, Johan Lindén, Georg Nöldeke, Ariel Rubinstein, and Benyamin Shitovitz for stimulating discussions.

** Department of Economics, University of Oslo, P.O.Box 1095 Blindern, N-0317 Oslo, Norway (Tel. +47-22855498, Fax. +47-22855035, Internet: geir.asheim@econ.uio.no).

*** CentER for Economic Research, Tilburg University, P.O.Box 90153, 5000 LE Tilburg, The Netherlands (Tel. +31-13-4663262, Fax. +31-13-4663066, Internet: martin.dufwenberg@kub.nl).

1. INTRODUCTION

Bernheim (1984) and Pearce (1984) assume that it is commonly known by the players in a normal form game that each player chooses a best response to a conjecture that does not assign positive probability to strategy vectors for his opponents known by the player to be impossible. This assumption implies that players choose strategies that survive iterated elimination of strongly dominated strategies.[1] Such strategies are called *rationalizable*. However, a rationalizable strategy can be weakly dominated by another (pure or mixed) strategy. It may seem natural to assume that players do not consider *inadmissible*, i.e. weakly dominated, strategies (see e.g. Luce & Raiffa (1957; Chapter 13) and Kohlberg & Mertens (1986) for supporting arguments). What are the implications of assuming that it is commonly known that players consider only admissible best responses? The present paper provides a new and, we claim, appropriate framework in which this question can be resolved.

Without further reflection one might be lead to conclude that a strategy is compatible with common knowledge of admissibility iff it survives iterated elimination of weakly dominated strategies (where at each round all inadmissible strategies are eliminated; this procedure will henceforth be referred to as *iterated admissibility*). To demonstrate why this conclusion is not obtained, and to motivate the subsequent discussion, consider two examples.

	L	***R***
U	1, 1	1, 1
M	0, 1	2, 0
D	1, 0	0, 1

G_1

In G_1,[2] iterated admissibility eliminates D in the first round, R in the second round, and M in the third round, so that U survives for player 1 and L survives for player 2. The

[1] This result holds if players are allowed to hold correlated conjectures concerning the choices of other players; see Tan & Werlang (1988). No independence restriction is imposed in the present paper. We follow e.g. Osborne & Rubinstein (1994) in using the term *rationalizable* even though correlated conjectures are allowed.

[2] This game is derived from an extensive game due to Battigalli (1989) and Börgers (1991).

procedure seems to force 2 into believing that *D* is less 'reasonable' than *M* for the sole reason that *D* is eliminated before *M*, even though both *M* and *D* are eventually judged to be 'unreasonable' choices for 1. This observation can be made precise by endowing each player with a *lexicographic probability system* (LPS; due to Blume, Brandenburger & Dekel (1991)) being a hierarchy of conjectures concerning the choices of his opponents. The player optimizes lexicographically given this hierarchy in the sense that he first optimizes at the highest level using the first-order conjecture and resolves any ties by using the second-order conjecture, etc. If every vector of strategies for the opponents is given positive probability by some conjecture in this hierarchy (i.e., the LPS has full support), such lexicographic optimization yields an admissible best response. Stahl (1993) requires that each player's LPS satisfy that, for each round *k*, any strategy eliminated by round *k* appears at a lower level in the hierarchy (hence, is deemed infinitely less likely) than *any* strategy not yet eliminated. He shows that a strategy survives iterated admissibility iff it is an admissible best response to a LPS satisfying this requirement.[3] This characterization means that iterated admissibility forces 2 to believe that *D* is infinitely less likely than *either* of *U* and *M* based on the premise that *D* is 1's only inadmissible strategy in G_1. On the basis of this premise such an inference seems at best to be questionable.

Brandenburger (1992) imposes a weaker restriction on each player's LPS, namely, for each round *k*, any strategy eliminated by round *k* appears at a lower level in the hierarchy (hence, is deemed infinitely less likely) than *some* strategy not yet eliminated. This means that 2 — based on the premise that only *D* is inadmissible — makes the unquestionable inference that *D* is infinitely less likely than *one* of *U* and *M*. Brandenburger assumes that it is commonly known that each player chooses an admissible best response to a full support LPS with a first-order conjecture that does not assign positive probability to strategy vectors for his opponents known by the player to be impossible. He shows that this assumption implies that players choose strategies that survive the Dekel-Fudenberg procedure[4] where one round of elimination

---

[3] A related result is established by Battigalli (1994). See also Rajan (1993) and Veronesi (1995).

[4] See Dekel & Fudenberg (1991). Börgers (1994) presents a similar foundation for the Dekel-Fudenberg procedure without using LPSs. While Brandenburger considers common *first-order* knowledge in the sense that each player in a lexicographic manner takes into account all strategy vectors for the opponents, Börgers considers *approximate* common knowledge in the sense that each player assigns positive probability to all strategy vectors for the opponents. See also Ben-Porath (1994) and Gul (1995).

of weakly dominated strategies is followed by iterated elimination of strongly dominated strategies. Brandenburger calls such strategies *permissible*. This procedure only eliminates *D* in the first round so that the permissible strategies are *U* and *M* for 1 and *L* and *R* for 2.

	L	***R***
U	1, 1	1, 1
M	1, 1	1, 0
D	0, 0	0, 1

G_2

While the Dekel-Fudenberg procedure seems to yield more reasonable implications than does iterated admissibility in G_1, this conclusion is reversed in G_2. Here iterated admissibility eliminates *D* in the first round and *R* in the second round (so that *U* and *M* survive for player 1 and *L* survives for player 2), while the Dekel-Fudenberg procedure only eliminates *D* in the first round (so that *U* and *M* survive for player 1 and *L* and *R* survive for player 2). Using the characterizations discussed above, the former procedure leads 2 to believe that *D* is infinitely less likely than *either* of *U* and *M*, while the latter procedure leads 2 to believe that *D* is infinitely less likely than *one* of *U* and *M*. In G_2, it seems natural to argue that 2 'should' believe that *D* is infinitely less likely than *either* of *U* and *M*. How can this intuition be captured without making the questionable inference that *D* is infinitely less likely than *either* of *U* and *M because D* is 1's only inadmissible strategy in G_2?

To address this issue, note that $\{U,M\}$ is 1's set of admissible best responses independently of what set of LPSs 1 may hold concerning the choices of 2. If 2 believes that any strategy outside 1's set of admissible best responses is infinitely less likely than any strategy contained in 1's set of admissible best responses, then 2 is lead to believe that *D* is infinitely less likely than *either* of *U* and *M*. This in turn implies that $\{L\}$ is 2's set of admissible best responses. The procedure just described eliminates strategy *sets*: In the first round, all strategy sets but $\{U,M\}$ are eliminated as 1's set of admissible best responses. In the second round, all strategy sets but $\{L\}$ are eliminated as 2's set of admissible best responses.

The present paper formalizes such a procedure — where strategy sets are iteratively eliminated — in terms of an increasing order of mutual knowledge. The paper's main contribution is to offer a consistent framework in which players can have knowledge of strategy sets for the opponents such that, for each opponent, the choice of any strategy outside her set is deemed infinitely less likely than the choice of any strategy contained in her set. The concept of a permissible strategy does not allow the players to have such knowledge. It may therefore, as illustrated by G_2, admit incautious behavior. The procedure of iterated admissibility effectively imposes such knowledge, but does not adequately explain how this is a consequence of an increasing order of mutual knowledge. G_1 illustrates how this can be problematic. The following framework is, thus, intended as a vehicle for amending these shortcomings.

Section 2 introduces a states-of-the-world model, with a partition of the state space for each player, and a function for each player assigning a subset of that player's strategies to each state. In Section 3 we assume that it is commonly known that each player's strategy set is the set of admissible best responses given a probability distribution that does not assign positive probability to vectors of strategy sets for his opponents known by the player to be impossible. If, in a two-player game, the probability distribution assigns all probability to a single strategy set for the opponent, then the player's set of admissible best responses is determined by considering only full support LPSs that satisfy the requirement that any strategy outside the opponent's set is deemed infinitely less likely than any strategy contained in the opponent's set. The set of admissible best responses is appropriately determined also with n players in the case of a non-degenerate probability distribution over vectors of strategy sets for the opponents. In this framework the assumption of common knowledge corresponds to a procedure where strategy sets are iteratively eliminated. A strategy set that survives the iterative elimination is called a *fully permissible set*.[5] For this concept, general existence is established, a characterization is offered, and a finite algorithm (that eliminates strategy sets) is provided.

---

[5] The term 'permissible' reflects — in line with Brandenburger's (1992) use — that it is commonly known that players perform lexicographic optimization (see our Definition 1). The term 'fully' reflects that a full support restriction is imposed, relative to subsets of the opponents' strategies (see our Definition 3). Note that the collection of sets that survive the elimination is independent of the order of elimination.

Section 4 presents properties of fully permissible sets and contains further examples. First, it is established that any strategy in a fully permissible set is both rationalizable and permissible (i.e., survives the Dekel-Fudenberg procedure). The converse does not hold since, in G_2, R is both a rationalizable and permissible strategy for 2, while $\{L\}$ is 2's unique fully permissible set in line with the above discussion.[6] Then, the "Battle-of-the-sexes-with-outside-option" and "Burning Money" games are used to show how the forward induction *outcome* results from an assumption of it being commonly known that each player's strategy set is the set of admissible best responses. In the latter example, burning need not be interpreted as a signal, implying that our analysis is robust against a critique commonly leveled at the usual forward induction *argument* in this game. Finally, a new perspective on the backward induction paradox is discussed in the context of the "Take-it-or-leave-it" game, in which there is for each player a fully permissible set that contains more than the backward induction strategy.

With reference to these examples it should be pointed out that our interpretation of normal form games differs from Pearce's (1984, p. 1031) who views these as "a convenient representation of a perfectly simultaneous game, in which no one can observe any move of any other player before moving himself". In contrast, we take a normal form game to represent *any* underlying extensive game. Building on recent results by Mailath, Samuelson & Swinkels (1993) we show that, once we insist on lexicographic optimization, sequential rationality is nevertheless adequately captured. Hence, the concept of fully permissible sets deals directly with both types of "imperfect" behavior discussed by Pearce: implausible behavior at unreached information sets as well as incautious optimization.

In Section 5 it is established how the present framework can, under alternative assumptions, be used to characterize rationalizable and permissible strategies through the concepts of *rationalizable* and *permissible sets.* The framework is also used to connect to the

---

[6] In G_2 $\{U,M\}$ is 1's unique fully permissible set. In G_1 all strategy sets but $\{U\}$, $\{M\}$ and $\{U,M\}$ are eliminated for 1 in the first round, $\{R\}$ is eliminated for 2 in the second round, and $\{M\}$ is eliminated for 1 in the third round. Hence, both $\{U\}$ and $\{U,M\}$ are fully permissible sets for 1 and both $\{L\}$ and $\{L,R\}$ are fully permissible sets for 2. The existence of multiple fully permissible sets means that common knowledge of each player's strategy set being the set of admissible best responses does not alone imply that each player gains knowledge of the strategy set of his opponent. Note that in G_1 — where one is lead to question the appropriateness of iterated admissibility — there is for each player a fully permissible set that contains more than the strategy surviving iterated admissibility.

analyses of Samuelson (1992) and Börgers & Samuelson (1992). However, we argue that, of the concepts considered, the concept of fully permissible sets captures common knowledge of admissibility in the most reasonable way. Section 6 presents an interpretation. All proofs are relegated to Appendix A, while Appendix B contains derivations for the examples.

2. THE FRAMEWORK

With $N = \{1, \ldots , n\}$ as the set of players, let S_i denote player i's finite set of pure strategies, and let $u_i\colon S \to \Re$ be i's payoff function, where $S = S_1 \times \ldots \times S_n = S_i \times S_{-i}$, and where $-i$ denotes $N \backslash \{i\}$. Then $G = (S, u)$ is a normal form game. Let $\Sigma_i := 2^{S_i} \backslash \{\emptyset\}$ denote the collection of non-empty subsets of S_i, and write $\Sigma = \Sigma_1 \times \ldots \times \Sigma_n = \Sigma_i \times \Sigma_{-i}$. We write p, r and s ($\in S$) for strategy vectors; P, R and X ($\subseteq S$) for subsets of strategy vectors; π, ρ and σ ($\in \Sigma$) for vectors of strategy sets; and Π, P and Ξ ($\subseteq \Sigma$) for subcollections of vectors of strategy sets. Note that if $\sigma = (\sigma_1, \ldots , \sigma_n) \in \Sigma$, then $\emptyset \neq \sigma_1 \times \ldots \times \sigma_n \subseteq S$. If $\emptyset \neq X_{-i} \subseteq S_{-i}$, let $\Delta(X_{-i})$ ($\Delta^0(X_{-i})$) denote the set of probability distributions on S_{-i} with support included in (equal to) X_{-i}, (with $\Delta(\cdot)$ and $\Delta^0(\cdot)$ later being used likewise for other finite sets). If $m_{-i} \in \Delta(S_{-i})$ is i's conjecture (subjective probability distribution) concerning the strategy choices of $-i$, abuse notation slightly by writing $u_i(s_i, m_{-i}(\cdot))$ for i's expected payoff given s_i and m_{-i}.

A *state* ω determines, for each player, a partition of the state space as well as a subset of that player's strategies. A state does not determine strategy choices. Formally, with Ω denoting the *state space,* each state $\omega \in \Omega$ specifies for each player i

- i's set of possible states given ω: $I_i(\omega) \subseteq \Omega$. The *information function* $I_i(\cdot)$ is *partitional* in the sense that there is a partition of Ω such that, $\forall \omega \in \Omega$, $I_i(\omega)$ is the element of the partition that contains ω. Write $K_iE := \{\omega' \in \Omega | \, I_i(\omega') \subseteq E\}$. Say that i *knows* the event $E \subseteq \Omega$ given ω if $\omega \in K_iE$. Since $\omega \in I_i(\omega)$, it follows that an event is true ($\omega \in E$) if i knows it.

- a non-empty subset of i's strategies: $\sigma_i(\omega) \in \Sigma_i$. Since $\forall\omega' \in I_i(\omega)$ are indistinguishable for i, we have that, $\forall\omega' \in I_i(\omega)$, $\sigma_i(\omega') = \sigma_i(\omega)$.

It follows that i knows given ω that the vector of the opponents' strategy sets is in $\Xi^i_{-i}(\omega) := \{\sigma_{-i}(\omega') \in \Sigma_{-i} | \ \omega' \in I_i(\omega)\}$. Note that, $\forall\omega \in \Omega$, $\sigma_{-i}(\omega) \in \Xi^i_{-i}(\omega)$ since $\omega \in I_i(\omega)$. Furthermore, $\Xi^i_{-i}(\omega') = \Xi^i_{-i}(\omega)$ if $\omega' \in I_i(\omega)$ since $\omega' \in I_i(\omega)$ implies $I_i(\omega') = I_i(\omega)$.

Write $KE := K_1E \cap ... \cap K_nE$. Say that the event $E \subseteq \Omega$ is *mutually known* given ω if $\omega \in KE$. Write $CKE := KE \cap KKE \cap KKKE ...$. Say that the event $E \subseteq \Omega$ is *commonly known* given ω if $\omega \in CKE$.

3. Fully permissible sets

The purpose of the present section is to introduce the admissible best response correspondence in the framework of Section 2 in order to model common knowledge of admissibility. The concept of a *lexicographic probability system* (LPS) is due to Blume et al. (1991). If $\forall k \in \{1, ..., K\}$, $m^k_{-i} \in \Delta(X_{-i})$, and $\bigcup_{k=1}^K \text{supp}(m^k_{-i}) = X_{-i}$, then $(m^1_{-i}, \dots, m^K_{-i})$ is a LPS with full support on X_{-i}. Following Veronesi (1995), let $\mathbf{L}\Delta^0(X_{-i})$ denote the set of LPSs with full support on X_{-i}.

DEFINITION 1. p_i is an *admissible best response* to $(m^1_{-i}, \dots, m^K_{-i}) \in \mathbf{L}\Delta^0(S_{-i})$ if, $\forall s_i \in S_i$, $(u_i(p_i, m^k_{-i}(\cdot)))^K_{k=1} \geq_L (u_i(s_i, m^k_{-i}(\cdot)))^K_{k=1}$.[7]

The following proposition establishes existence and provides characterizations.

PROPOSITION 1. *(i) For any* $(m^1_{-i}, \dots, m^K_{-i}) \in \mathbf{L}\Delta^0(S_{-i})$ *there exists an admissible best response to* $(m^1_{-i}, \dots, m^K_{-i})$. *(ii) There exists* $(m^1_{-i}, \dots, m^K_{-i}) \in \mathbf{L}\Delta^0(S_{-i})$ *such that* p_i *is an admissible best response to* $(m^1_{-i}, \dots, m^K_{-i})$ *iff* p_i *is not weakly dominated (by a pure or mixed strategy). (iii) Let* Γ *be an extensive game without nature, with* G *being the corresponding pure strategy reduced normal form.*[8] *If* p_i *is an admissible best response to* $(m^1_{-i}, \dots, m^K_{-i}) \in \mathbf{L}\Delta^0(S_{-i})$, *then*

---

[7] For two vectors v and w, $v \geq_L w$ iff whenever $w_k > v_k$, there exists $\ell < k$ such that $v_\ell > w_\ell$.

[8] See Mailath et al. (1993, Def. 1).

$(m^1_{-i},\ldots,m^K_{-i})$ *generates in* Γ *a system of conjectures satisfying Bayes' law such that* p_i *maximizes expected payoff at all of i's information sets that can be reached given* p_i.

Given $\omega \in \Omega$, endow *i* with a subjective probability distribution μ_{-i} over the vectors of opponents' strategy sets in $\Xi^i_{-i}(\omega)$. To determine *i*'s set of admissible best responses given μ_{-i}, it is necessary to specify how μ_{-i} restricts the set of LPSs that *i* can hold. For simplicity and to enable comparison with existing literature, we choose a formulation where μ_{-i} imposes a requirement on first-order conjectures only. Definition 2 below is a necessary requirement, which — being insufficient for our purposes — is subsequently strengthened through Definition 3.

If *i* knows that the opponents' vector is $\sigma_{-i} = (\sigma_j)_{j\neq i}$ — i.e., $\Xi^i_{-i}(\omega) = \{\sigma_{-i}\}$ — so that μ_{-i} is degenerate, then $(m^1_{-i},\ldots,m^K_{-i}) \in \mathbf{L}\Delta^0(S_{-i})$ must not assign positive first-order probability to any strategy vector outside $\times_{j\neq i}\sigma_j$; i.e. $m^1_{-i} \in \Delta(\times_{j\neq i}\sigma_j)$. When μ_{-i} is not degenerate, this translates into the requirement that the first-order conjecture be consistent with μ_{-i}.

DEFINITION 2. $m_{-i} \in \Delta(S_{-i})$ is *consistent* with $\mu_{-i} \in \Delta(\Sigma_{-i})$ if, $\forall\sigma_{-i} = (\sigma_j)_{j\neq i} \in \Sigma_{-i}$, $\exists m^{\sigma_{-i}}_{-i} \in \Delta(\times_{j\neq i}\sigma_j)$ such that, $\forall s_{-i} \in S_{-i}$, $m_{-i}(s_{-i}) = \sum_{\sigma_{-i}\in\Sigma_{-i}} \mu_{-i}(\sigma_{-i}) m^{\sigma_{-i}}_{-i}(s_{-i})$.[9]

Returning to the case of a degenerate μ_{-i}, a next step is to require that $(m^1_{-i},\ldots,m^K_{-i}) \in \mathbf{L}\Delta^0(S_{-i})$ assigns positive first-order probability to any strategy vector contained in $\times_{j\neq i}\sigma_j$; i.e. $m^1_{-i} \in \Delta^0(\times_{j\neq i}\sigma_j)$ or, equivalently, $(m^1_{-i}) \in \mathbf{L}\Delta^0(\times_{j\neq i}\sigma_j)$. This is in line with Pearce's (1984) formulation of cautiousness and ensures that any strategy vector outside $\times_{j\neq i}\sigma_j$ is deemed infinitely less likely than any strategy vector contained in $\times_{j\neq i}\sigma_j$. When μ_{-i} is not degenerate, this translates into the requirement that the first-order conjecture be fully consistent with μ_{-i}.

DEFINITION 3. $m_{-i} \in \Delta(S_{-i})$ is *fully consistent* with $\mu_{-i} \in \Delta(\Sigma_{-i})$ if, $\forall\sigma_{-i} = (\sigma_j)_{j\neq i} \in \Sigma_{-i}$, $\exists m^{\sigma_{-i}}_{-i} \in \Delta^0(\times_{j\neq i}\sigma_j)$ such that, $\forall s_{-i} \in S_{-i}$, $m_{-i}(s_{-i}) = \sum_{\sigma_{-i}\in\Sigma_{-i}} \mu_{-i}(\sigma_{-i}) m^{\sigma_{-i}}_{-i}(s_{-i})$.

When μ_{-i} is degenerate, it appears more general to allow for multiple levels of conjectures inside $\times_{j\neq i}\sigma_j$ by having $(m^1_{-i},\ldots,m^k_{-i}) \in \mathbf{L}\Delta^0(\times_{j\neq i}\sigma_j)$ for *some* $k \in \{1, \ldots, K\}$. By Pearce

---

[9] Let ϕ_{-i} be defined by, for all $\sigma_{-i} \in \Sigma_{-i}$, $\phi_{-i}(\sigma_{-i}) = \sum_{\emptyset\neq\xi_{-i}\subseteq\sigma_{-i}} \mu_{-i}(\xi_{-i})$. In the terminology of Hendon et al. (1994), μ_{-i} is a *mass* function, ϕ_{-i} is a *belief* function, and Definition 2 requires m_{-i} to be in the *core* of ϕ_{-i}.

(1984, Lemma 4), imposing $k = 1$ is *not* too restrictive since — in the game derived from G by restricting the opponents' set of strategy vectors to $\times_{j \neq i} \sigma_j$ — a first-order full support constraint precludes a strategy for i iff it is weakly dominated. In order to preclude any weakly dominated strategy in the unrestricted G, i must, however, perform lexicographic optimization. Since the first-order conjecture satisfies the relevant full support constraint, no restriction other than $\bigcup_{k=1}^{K} \text{supp}(m_{-i}^k) = S_{-i}$ need be imposed on lower level conjectures. This holds independently of whether μ_{-i} is degenerate or not. Therefore, say that $(m_{-i}^1, \ldots, m_{-i}^K) \in \mathbf{L}\Delta^0(S_{-i})$ is *fully first-order consistent* with $\mu_{-i} \in \Delta(\Sigma_{-i})$ if m_{-i}^1 is *fully consistent* with μ_{-i}. Full first-order consistency implies, with $n = 2$, that any strategy *not* in an opponent set that is assigned positive probability by μ_{-i}, is deemed infinitely less likely than any strategy in such a set.

If $\mu_{-i} \in \Delta(\Sigma_{-i})$, let

$$a_i^0(\mu_{-i}) := \{p_i \in S_i \mid \exists (m_{-i}^1, \ldots, m_{-i}^K) \text{ fully first-order consistent with } \mu_{-i}$$

$$\text{such that } p_i \text{ is an admissible best response to } (m_{-i}^1, \ldots, m_{-i}^K)\}.$$

By Proposition 1(i), $a_i^0(\mu_{-i})$ is a non-empty strategy set for i. If $\emptyset \neq \Xi_{-i} \subseteq \Sigma_{-i}$, let

$$\alpha_i^0(\Xi_{-i}) := \{\, a_i^0(\mu_{-i}) \mid \mu_{-i} \in \Delta(\Xi_{-i}) \,\}$$

denote the collection that contains a strategy set iff it is the set of admissible best responses to LPSs that are fully first-order consistent with some probability distribution in $\Delta(\Xi_{-i})$. If $\emptyset \neq \Xi'_{-i} \subseteq \Xi''_{-i} \subseteq \Sigma_{-i}$, then $\emptyset \neq \alpha_i^0(\Xi'_{-i}) \subseteq \alpha_i^0(\Xi''_{-i}) \subseteq \Sigma_i$. If $\emptyset \neq \Xi = \Xi_1 \times \ldots \times \Xi_n \subseteq \Sigma$, write $\alpha^0(\Xi) := \alpha_1^0(\Xi_{-1}) \times \ldots \times \alpha_n^0(\Xi_{-n})$. Let $A_i^0 := \{\omega \in \Omega \mid \sigma_i(\omega) \in \alpha_i^0(\Xi_{-i}^i(\omega))\}$, with $A^0 := A_1^0 \cap \ldots \cap A_n^0$. If $\omega \in A^0$, then, $\forall i \in N$, $\sigma_i(\omega)$ is the set of admissible best responses to LPSs that are fully first-order consistent with a probability distribution that does not assign positive probability to vectors of opponents' sets known by i, given ω, to be impossible. The concept of fully permissible sets can now be defined and characterized.

DEFINITION 4. A non-empty strategy set π_i is a *fully permissible set* for i if there exists $\omega \in CKA^0$ with $\sigma_i(\omega) = \pi_i$.

PROPOSITION 2. *A non-empty strategy set* π_i *is a fully permissible set for i iff there exists* $\Xi = \Xi_1 \times \ldots \times \Xi_n$ *with* $\pi_i \in \Xi_i$ *such that* $\Xi \subseteq \alpha^0(\Xi)$.

The following proposition establishes general existence and provides an algorithm.

PROPOSITION 3. *Let* $\Pi^0 = \Pi_1^0 \times \ldots \times \Pi_n^0$ *denote the collection of vectors of fully permissible sets. (i)* $\forall i \in N$, $\Pi_i^0 \neq \emptyset$. *(ii)* $\Pi^0 = \alpha^0(\Pi^0)$. *(iii) The sequence defined by* $\Xi(0) = \Sigma$ *and,* $\forall k \geq 1$, $\Xi(k) = \alpha^0(\Xi(k-1))$ *converges to* Π^0 *in a finite number of iterations.*

Note that $\Pi^0 = \alpha^0(\Pi^0)$ means that Π^0 is a fixed point in terms of a collection of vectors of strategy sets. By Proposition 2 it is the largest such fixed point. The algorithm of Proposition 3(iii) has the usual interpretation in terms of an increasing order of mutual knowledge.

4. PROPERTIES AND EXAMPLES

By the following proposition, the concept of fully permissible sets refines rationalizability (as defined by Bernheim (1984) and Pearce (1994) except that we here consider pure strategies only and allow conjectures to be correlated). G_2 of the introduction as well as G_3 and G_4 below illustrate that this refinement can be strict.

PROPOSITION 4. *Let* $R = R_1 \times \ldots \times R_n$ *denote the set of rationalizable strategy vectors; i.e. strategy vectors surviving iterated elimination of strongly dominated strategies. Then,* $\forall i \in N$, $r_i \in R_i$ *if there exists a fully permissible set* π_i *for i such that* $r_i \in \pi_i$.

The following proposition establishes that any strategy in a fully permissible set survives the Dekel-Fudenberg (1991) procedure. Again, G_2 as well as G_3 and G_4 illustrate that not all strategies surviving the Dekel-Fudenberg procedure are elements of some fully permissible set.

PROPOSITION 5. *Let* $P = P_1 \times \ldots \times P_n$ *denote the set of permissible strategy vectors; i.e. strategy vectors surviving one round of elimination of weakly dominated strategies and then* *iterated elimination of strongly dominated strategies. Then,* $\forall i \in N$, $p_i \in P_i$ *if there exists a fully permissible set* π_i *for i such that* $p_i \in \pi_i$.

Below G_3 and G_4 illustrate a notion of forward induction, while G_5 is included to discuss backward induction. Each of these games is interpreted as the pure strategy reduced normal form (PRNF) of an extensive game. The foundation for analyzing these games in the PRNF is given in Proposition 1(iii), which in turn is based on results by Mailath et al. (1993).

G_3 is the PRNF of a "Battle-of-the-Sexes-with-an-outside-option" game, where 1 and 2 move in sequence, with 2 being asked to play only if 1 does not choose the outside option *U*. Such a game, first introduced by Kreps & Wilson (1982) (who credit Elon Kohlberg), has been widely used to illustrate *forward induction.* Pearce (1984) uses it to promote his extensive form rationalizability. Kohlberg & Mertens (1986) argue that the information contained in the PRNF G_3 should suffice to analyze any underlying extensive game.

	L	***R***
U	2, 2	2, 2
M	3, 1	0, 0
D	0, 0	1, 3

G_3

Within the states-of-the-world model of Section 2, assume — as specified in Section 3 — that it is commonly known that each player's strategy set is the set of admissible best responses to LPSs that are fully consistent with some probability distribution that does not assign positive probability to strategy sets for his opponent known by the player to be impossible. Since *D* is a strongly dominated strategy, *D* cannot be an element of 1's set of admissible best responses. This *does not* imply — as is implicitly the case in the procedure of iterated admissibility (see Stahl (1993)) — that 2 knows that *D* is infinitely less likely than *M*. However, 2 knows that only $\{U\}$, $\{M\}$ and $\{U,M\}$ are possible candidates for 1's strategy set. This excludes $\{R\}$ as 2's strategy set, since $\{R\}$ is 2's set of admissible best responses only if 2 assigns positive probability to $\{D\}$ or $\{U,D\}$ being 1's strategy set. This in turn means that 1 knows that only $\{L\}$ and $\{L,R\}$ are possible candidates for 2's strategy set, implying that $\{U\}$

cannot be 1's set of admissible best responses. Knowing that only $\{M\}$ or $\{U,M\}$ are candidates for 1's strategy set *does* imply that 2 knows that *D* is infinitely less likely than *M*. Hence, 2's set of admissible best responses is $\{L\}$, and, therefore, 1's set of admissible best responses is $\{M\}$. The argument above shows that $(\{M\},\{L\})$ is the unique vector of fully permissible sets. The strategy profile implied by this vector entails that 1 can signal — by asking 2 to play — that he seeks a payoff of at least 2, leading to the implementation of 1's preferred outcome. In the procedure of iterated admissibility such signaling hinges solely on the fact that *M* is an admissible strategy for 1 while *D* is not. In contrast, it here relies on the properties of the entire game. This difference is reflected by the fact that, while iterated admissibility converges after only *three* rounds of elimination of strategies, the present procedure needs *five* rounds of elimination of strategy sets; i.e. at least 4th order of mutual knowledge is required.

Turn now to the "Burning Money" game due to van Damme (1989, Fig. 5) and Ben-Porath & Dekel (1992, Fig. 1.2). G_4 is the PRNF of a "Battle-of-the-Sexes" (B-o-S) game with the additional feature that 1 can publicly destroy 1 unit of payoff before the B-o-S game starts. *BU* (*NU*) is the strategy where 1 burns (does not burn), and then plays *U*, etc., while *LR* is the strategy where 2 responds with *L* conditional on 1 not burning and *R* conditional on 1 burning, etc. The forward induction outcome (supported e.g. by iterated admissibility) involves implementation of 1's preferred B-o-S outcome, with *no payoff being burnt*. One might be skeptical to the use of iterated admissibility in the "Burning Money" game because it effectively requires 2 to believe that *BD* is infinitely less likely than *BU* although all strategies involving burning (i.e. both *BD* and *BU*) are eventually eliminated by the procedure. As demonstrated in Appendix B, common knowledge of admissibility in the sense of Definition 4 corresponds to an iterative procedure, where at no stage of the iteration need 2 believe that *BD* is infinitely less likely than *BU* since $\{NU\}$ is always included as a possible strategy set for 1. The procedure uniquely determines $\{NU\}$ as 1's fully permissible set and $\{LL,LR\}$ as 2's fully permissible set.[10] Even though the forward induction *outcome* is obtained, 2 is free to have any conjecture

[10] Also Battigalli (1989), Asheim (1994), and Dufwenberg (1994) argue that (NU,LR) in addition to (NU,LL) is a viable strategy vector in "Burning Money".

conditional on burning; hence, she need not interpret burning as a signal that 1 will play according to his preferred B-o-S outcome.

	LL	***LR***	***RL***	***RR***
NU	3, 1	3, 1	0, 0	0, 0
ND	0, 0	0, 0	1, 3	1, 3
BU	2, 1	-1, 0	2, 1	-1, 0
BD	-1, 0	0, 3	-1, 0	0, 3

G_4

It is noteworthy that common knowledge of admissibility in the sense of Definition 4 yields the forward induction outcomes in G_3 and G_4. In addition to iterated admissibility and the equivalent procedure suggested by Stahl (1993), Pearce's (1984) extensive form rationalizability — both as originally defined and as characterized by Battigalli (1995) — yields the forward induction outcomes in G_3 and G_4. In contrast to the present analysis, these procedures have not formally been given a common knowledge basis, unless a very strong primitive assumption like Stahl's (1993) 'iterated lexicographic coherence' is accepted.

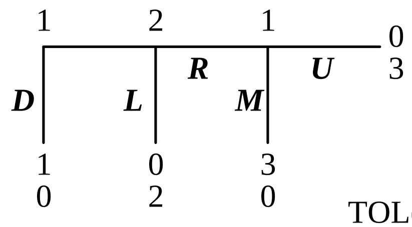
In a perfect information extensive game with generic payoffs, both procedures of the previous paragraph lead to the backward induction outcome. During the last few years, a number of papers have discussed whether backward induction follows from an assumption that it is commonly known that players are rational in the sense of maximizing expected payoff at all decision nodes.[11] The background for this interest is the following paradoxical aspect of backward induction: Why should a player believe that an opponent's future play will satisfy backward induction if the opponent's previous play is incompatible with backward induction?

Reny (1993) studies the "Take-it-Or-Leave-it" game with k stages (TOL(k)), where at the ℓth stage of the game, the total pot is ℓ dollars. If ℓ is odd (even), player 1 (2) may take the ℓ dollars and end the game, or leave it, in which case the pot increases with one dollar.

---

[11] These papers include Aumann (1995), Basu (1990), Ben-Porath (1994), Bicchieri (1989), Binmore (1987, 1995), Gul (1995), and Reny (1993).

Should the game continue until the kth stage and the player whose turn it is decides to leave the k dollars, it is given to the other player. We analyze TOL(3) in detail.



TOL(3)

	L	R
U	0, 2	0, 3
M	0, 2	3, 0
D	1, 0	1, 0

G_5

G_5 is the PRNF of TOL(3). Since U is a strongly dominated strategy, U cannot be an element of 1's set of admissible best responses. Hence, 2 knows that only $\{M\}$, $\{D\}$ and $\{M,D\}$ are possible candidates for 1's strategy set. This excludes $\{R\}$ as 2's strategy set, since $\{R\}$ is 2's set of admissible best responses only if 2 assigns positive probability to $\{U\}$ or $\{U,D\}$. This in turn means that 1 knows that only $\{L\}$ and $\{L,R\}$ are possible candidates for 2's strategy set, implying that $\{M\}$ cannot be 1's set of admissible best responses. No further elimination of strategy sets is possible. Hence, 1's collection of fully permissible sets is $\{\{D\}, \{M,D\}\}$, and 2's collection of fully permissible sets is $\{\{L\}, \{L,R\}\}$. For each player, the smaller set contains only the backward induction strategy, while the larger set coincides with the set of strategies surviving the Dekel-Fudenberg procedure. Before discussing, in the context of TOL(3), the perspective offered by the concept of fully permissible sets on the backward induction paradox, note that the fully permissible sets justify each other as follows:

$$\alpha_1^0(\mu_{-1}) = \{D\} \quad \text{if } \mu_{-1}(\{L\}) \geq {}^2/_3 \text{ and } \mu_{-1}(\{L,R\}) = 1 - \mu_2(\{L\})$$

$$\alpha_1^0(\mu_{-1}) = \{M,D\} \quad \text{if } \mu_{-1}(\{L\}) < {}^2/_3 \text{ and } \mu_{-1}(\{L,R\}) = 1 - \mu_2(\{L\})$$

$$\alpha_2^0(\mu_{-2}) = \{L\} \quad \text{if } \mu_{-2}(\{D\}) < 1 \text{ and } \mu_{-2}(\{M,D\}) = 1 - \mu_{-2}(\{D\})$$

$$\alpha_2^0(\mu_{-2}) = \{L,R\} \quad \text{if } \mu_{-2}(\{D\}) = 1.$$

With common knowledge of admissibility in the sense of Definition 4, 2 knows that all sets for 1 but $\{D\}$ and $\{M,D\}$ are impossible. This implies that 2 must deem the choice of U infinitely less likely than the choice of D; it does *not* imply that 2 knows that some strategy choice — not even U — is impossible. Hence, 2 cannot observe an impossible strategy choice.

Consider 2's conjecture concerning 1's strategy choice conditional on 2 being asked to play, i.e., conditional on 1 choosing *U* or *M*. If 2 assigns positive probability to {*M*,*D*}, then 2's first-order conjecture assigns zero probability to *U* and positive probability to *M*. Hence, 2's conditional conjecture assigns probability 1 to 1 choosing *M*, implying that only *L* is an admissible best response. If 2 assigns probability 1 to {*D*}, then 2's first-order conjecture assigns zero probability to both *U* and *M*. Since being asked to play is a zero probability (but not impossible!) event that can be caused by 1 choosing outside his strategy set, there are no restrictions on 2's conditional conjecture. Therefore, 2 may — conditional on 1 choosing *U* or *M* — assign positive probability to 1 choosing the strongly dominated strategy *U*, implying that *R* in addition to *L* is an admissible best response given that 2 assigns probability 1 to {*D*}.

Ben-Porath (1994) agrees with the present analysis by not having players know that particular strategy choices are impossible. Rather, players can be *certain* (in the sense of believing with probability 1) at the beginning of the game that particular strategies will not be chosen. He assumes that there is, at the beginning of the game, common certainty that each player is rational in the sense of maximizing expected payoff at all the player's decision nodes. This entails that a player cannot assign positive unconditional probability to a strategy vector for the opponents which contradicts common certainty of rationality. However, for a given unconditional conjecture Ben-Porath imposes only Bayes' law on conditional conjectures. This is analogous to Brandenburger's (1992) support restriction on first-order conjectures only. Indeed, Ben-Porath shows that common certainty of rationality corresponds to strategy vectors surviving the Dekel-Fudenberg procedure. In TOL(3) his approach implies that if 2 assigns probability 1 to 1 choosing *D*, then 2, if asked to play, is free to have any conditional conjecture concerning 1's strategy choice since Bayes' law does not apply. Hence, 2 may conditionally assign positive probability to *U*, thereby allowing *R* to maximize expected payoff.

Aumann's (1995) analysis of common knowledge of rationality in perfect information games requires fully specified strategies; i.e., for each player, actions must also be specified at nodes that the player's own strategy precludes from being reached. Hence, the PRNF — implying that in the extensive form only *plans of actions* (Rubinstein (1991)) are determined — is not sufficient. With Aumann's assumptions, common knowledge of rationality implies in

TOL(3) that all strategies for 1 but the one where he takes the 1 dollar at his first node and takes the 3 dollars at his last node are impossible. Hence, it is impossible for 2 to be asked to play. However, in the *counterfactual* event that 2 is asked to play, she acts as if 1 at his last node follows his only possible strategy, implying that it is impossible for 2 to choose *R*. Thus, in Aumann's analysis, if common knowledge of rationality obtains, then each player chooses the backward induction strategy. The present analysis is based on the argument that only plans of actions matter in a context where players do not make mistakes and conjectures concerning the choices of opponents are explicitly specified. That 2 is asked to play is seen to be incompatible with 1 having planned to take to the 1 dollar at his first node.

In summary, Aumann (1995) assumes that players know that particular strategies are impossible. This allows for a common knowledge formulation since actual play not in accordance with common knowledge of rationality cannot occur. Ben-Porath (1994) does not let players know that particular strategies are impossible; knowledge must therefore be weakened to certainty. Since actual play not in accordance with common certainty of rationality can occur, in general, common certainty of rationality holds only at the beginning of the game. The present analysis allows for a common knowledge formulation by letting players have knowledge of strategy sets. Players do not know that particular strategies are impossible; still, common knowledge of admissibility cannot be contradicted by actual play.

5. Alternative formulations

The concept of fully permissible sets (Definition 4) relies on a full support restriction on first-order conjectures (Definition 3) and on lexicographic optimization (Definition 1). The present section demonstrates that the chosen formulation is tight in the sense that incautious behavior cannot be ruled out if the full support restriction is relaxed, or if the players do not perform lexicographic optimization. As a byproduct, this exercise allows us to establish important connections to earlier contributions.

First, consider the consequences of relaxing the full support restriction on first-order conjectures. Therefore return to Definition 2, and say that $(m^1_{-i},\ldots,m^K_{-i}) \in \mathbf{L}\Delta^0(S_{-i})$ is *first-order consistent* with $\mu_{-i} \in \Delta(\Sigma_{-i})$ if m^1_{-i} is *consistent* with μ_{-i}. First-order consistency implies, with $n = 2$, that the support of the first-order conjecture is included in the union of sets that are assigned positive probability by μ_{-i}. If $\mu_{-i} \in \Delta(\Sigma_{-i})$, let $a_i(\mu_{-i}) := \{p_i \in S_i| \exists$ $(m^1_{-i},\ldots,m^K_{-i})$ first-order consistent with μ_{-i} such that p_i is an admissible best response to $(m^1_{-i},\ldots,m^K_{-i})\}$, and if $\emptyset \neq \Xi_{-i} \subseteq \Sigma_{-i}$, let $\alpha_i(\Xi_{-i}) := \{a_i(\mu_{-i})|\ \mu_{-i} \in \Delta(\Xi_{-i})\}$. Let $A_i := \{\omega \in \Omega|\ \sigma_i(\omega) \in \alpha_i(\Xi^i_{-i}(\omega))\}$, with $A := A_1 \cap \ldots \cap A_n$. The concept of permissible sets can now be defined. The analogues of Propositions 2 and 3 are available.

DEFINITION 5. A non-empty strategy set π_i is a *permissible set* for i if there exists $\omega \in CKA$ with $\sigma_i(\omega) = \pi_i$.

PROPOSITION 6. *Let* $P = P_1 \times \ldots \times P_n$ *denote the set of permissible strategy vectors; i.e. strategy vectors surviving one round of elimination of weakly dominated strategies and then iterated elimination of strongly dominated strategies. Then,* $\forall i \in N$, $p_i \in P_i$ *iff there exists a permissible set* π_i *for i such that* $p_i \in \pi_i$.

Hence, the set of i's permissible strategies equals the union of i's permissible sets. This means that permissible strategies — as defined by Brandenburger (1992) so that a strategy is permissible iff it survives the Dekel-Fudenberg procedure — can be characterized by the states-of-the-world model of Section 2 even though a state, for each player, determines a strategy set rather than a strategy. In G_2 of the introduction, a choice of the permissible strategy R seems to entail incautious behavior. This strategy is not contained in any fully permissible set, but is — according to Proposition 6 — contained in some permissible set.

Secondly, consider the consequences of relaxing the assumption that players perform lexicographic optimization. Instead, assume ordinary optimization.

DEFINITION 6. r_i is a *best response* to $m_{-i} \in \Delta(S_{-i})$ if, $\forall s_i \in S_i$, $u_i(r_i, m_{-i}(\cdot)) \geq u_i(s_i, m_{-i}(\cdot))$.

By the finiteness of G, it follows that, for any $m_{-i} \in \Delta(S_{-i})$, there exists a best response to m_{-i}.

We first combine the best response correspondence with Definition 2. If $\mu_{-i} \in \Delta(\Sigma_{-i})$, let $b_i(\mu_{-i}) := \{r_i \in S_i | \exists m_{-i}$ consistent with μ_{-i} such that r_i is a best response to $m_{-i}\}$, and if $\emptyset \neq \Xi_{-i} \subseteq \Sigma_{-i}$, let $\beta_i(\Xi_{-i}) := \{ b_i(\mu_{-i}) \,|\, \mu_{-i} \in \Delta(\Xi_{-i}) \}$. Let $B_i := \{\omega \in \Omega | \sigma_i(\omega) \in \beta_i(\Xi^i_{-i}(\omega)) \}$, with $B := B_1 \cap ... \cap B_n$. The concept of rationalizable sets can now be defined. The analogues of Propositions 2 and 3 are available.

DEFINITION 7. A non-empty strategy set ρ_i is a *rationalizable set* for *i* if there exists $\omega \in CKB$ with $\sigma_i(\omega) = \rho_i$.

PROPOSITION 7. *Let* $R = R_1 \times ... \times R_n$ *denote the set of rationalizable strategy vectors; i.e. strategy vectors surviving iterated elimination of strongly dominated strategies. Then,* $\forall i \in N$, $r_i \in R_i$ *iff there exists a rationalizable set* ρ_i *for i such that* $r_i \in \rho_i$.

Hence, the set of *i*'s rationalizable strategies equals the union of *i*'s rationalizable sets. This means that rationalizability — as defined by Bernheim (1984) and Pearce (1984) but allowing for correlated conjectures so that a strategy is rationalizable iff it survives iterated strong dominance — can be characterized by the states-of-the-world model of Section 2 even though a state, for each player, determines a strategy set rather than a strategy. There may exist multiple rationalizable sets for each player; games with multiple strict Nash equilibria illustrates this since any strict Nash equilibrium constitutes a vector of rationalizable sets.

We then combine the best response correspondence with Definition 3, thus imposing a full support constraint of the kind considered by Samuelson (1992) and Börgers & Samuelson (1992). If $\mu_{-i} \in \Delta(\Sigma_{-i})$, let $b^0_i(\mu_{-i}) := \{r_i \in S_i | \exists m_{-i}$ fully consistent with μ_{-i} such that r_i is a best response to $p_{-i}\}$, and if $\emptyset \neq \Xi_{-i} \subseteq \Sigma_{-i}$, let $\beta^0_i(\Xi_{-i}) := \{ b^0_i(\mu_{-i}) \,|\, \mu_{-i} \in \Delta(\Xi_{-i}) \}$. Let $B^0_i := \{\omega \in \Omega | \sigma_i(\omega) \in \beta^0_i(\Xi^i_{-i}(\omega)) \}$, with $B^0 := B^0_1 \cap ... \cap B^0_n$. The concept of fully rationalizable sets can now be defined. The analogues of Propositions 2 and 3 are available.

DEFINITION 8. A non-empty strategy set ρ_i is a *fully rationalizable set* for *i* if there exists $\omega \in CKB^0$ with $\sigma_i(\omega) = \rho_i$.

In order to show that rationalizable and fully rationalizable sets may admit incautious behavior, consider G_6 due to Samuelson (1992, Ex. 8) and Börgers & Samuelson (1992, Ex. 3). Here, the inadmissible strategy D is contained in some fully rationalizable set since the collection of vectors of fully permissible sets is $\{\{U\}, \{U,D\}\}\times\{\{L\}, \{L,R\}\}$. It is straightforward to show that any strategy in a fully rationalizable set is rationalizable; hence, it follows that D is also rationalizable. A strategy in a fully rationalizable set need not be permissible; G_6 illustrates this since only U and L are permissible strategies. By Proposition 3(i) and Proposition 5, $\{\{U\}\}\times\{\{L\}\}$ is the collection of vectors of fully permissible sets.

	L	***R***
U	1, 1	1, 0
D	1, 0	0, 1

G_6

Our terminology and results are summarized in the following table.

	Optimization	Lexicographic optimization
Support restriction on first-order conjectures	*rationalizable sets* characterizes rationalizable strategies	*permissible sets* refines rationalizable strat. charact. permissible strat.
Full support restriction on first-order conjectures	*fully rationalizable sets* refines rationalizable strategies	*fully permissible sets* refines rationalizable strat. refines permissible strat.

TABLE 1

We have argued that — of the concepts included in Table 1 — the concept of fully permissible sets captures common knowledge of admissibility in the most reasonable way.

6. A COMMENT ON INTERPRETATION

We conclude by suggesting a possible intuitive interpretation of the states-of-the-world model of Section 2 as specified in Section 3. Consider a structure where

- the players receive private recommendations from separate analysts, where the analysts may coincide with the players themselves,
- each analyst's recommendation is given in the form of a strategy set,
- for each player, the choice of any strategy outside his set of recommendation is deemed infinitely less likely than the choice of any strategy contained in his set of recommendation,
- each analyst offers a recommendation that is the set of admissible best responses given the analyst's subjective probability distribution over vectors of strategy sets that the analyst thinks are possible sets of recommendation for the opponents.

A fully permissible set is a possible set of recommendation when this structure is commonly known by the analysts. In line with the discussion in Aumann & Brandenburger (1995, pp. 1174–1175), the states-of-the-world model is *descriptive* when referring to the recommendations of the analysts. However, given the suggested interpretation it seems appropriate to say that the model *prescriptive* when referring to the strategy choices of the players.

APPENDIX A: PROOFS

Proof of Prop. 1. Given $(m_{-i}^1,\dots,m_{-i}^K)$, write $Z_i^0 := S_i$ and define Z_i^1, Z_i^2, ... inductively by $Z_i^k := \arg\max_{s_i \in Z_i^{k-1}} u_i(s_i, m_{-i}^k(\cdot))$ for $k \in \{1,\dots,K\}$. Then p_i is an admissible best response to $(m_{-i}^1,\dots,m_{-i}^K)$ iff $p_i \in Z_i^K$. *(i)* By the finiteness of G, $Z_i^K \neq \emptyset$. *(ii)* *(If)* By Pearce (1984, Lemma 4), if p_i is not weakly dominated by a pure or mixed strategy, there exists $m_{-i} \in \Delta^0(S_{-i})$ such that, $\forall s_i \in S_i$, $u_i(p_i, m_{-i}(\cdot)) \geq u_i(s_i, m_{-i}(\cdot))$. Note that $(m_{-i}) \in \mathbf{L}\Delta^0(S_{-i})$. *(Only if)* Assume that p_i is weakly dominated by a (possibly degenerate) mixed strategy $m_i \in \Delta(S_i)$. It suffices to show that, $\forall (m_{-i}^1,\dots,m_{-i}^K) \in \mathbf{L}\Delta^0(S_{-i})$, $p_i \notin Z_i^K$. Note that $\{p_i\} \cup \text{supp}(m_i) \subseteq Z_i^0$. Furthermore,

$\forall k \in \{1,...,K\}$, $\{p_i\} \cup \text{supp}(m_i) \subseteq Z_i^{k-1}$ implies ($p_i \in Z_i^k$ only if $\text{supp}(m_i) \subseteq Z_i^k$). To see this, observe that $u_i(p_i, m_{-i}^k(\cdot)) \leq \sum_{s_i \in S_i} m_i(s_i) u_i(s_i, m_{-i}^k(\cdot))$ since m_i weakly dominates p_i. Also, if $p_i \in Z_i^k$, $\forall r_i \in Z_i^{k-1}$, $u_i(p_i, m_{-i}^k(\cdot)) \geq u_i(r_i, m_{-i}^k(\cdot))$. Hence, $p_i \in Z_i^k$ implies, $\forall s_i \in \text{supp}(m_i) \subseteq Z_i^{k-1}$, $\forall r_i \in Z_i^{k-1}$, $u_i(p_i, m_{-i}^k(\cdot)) = u_i(s_i, m_{-i}^k(\cdot)) \geq u_i(r_i, p_{-i}^k(\cdot))$. However, $\{p_i\} \cup \text{supp}(m_i) \subseteq Z_i^K$ contradicts that m_i weakly dominates p_i. *(iii)* For each information set h for i in Γ, there exists a corresponding set $S(h) \subseteq S$ in G; see Mailath et al. (1993, Section 2). By perfect recall, $S(h) = S_i(h) \times S_{-i}(h)$. Write $H_i := \{S(h) \subseteq S |\ h$ is an information set for i in $\Gamma\}$ and $H_i(s_i) := \{X \in H_i |\ s_i \in X_i\}$. If $X_i \times X_{-i} \in H_i$, let $m_{-i}^{X_{-i}} \in \Delta(X_{-i})$ be defined by, $\forall s_{-i} \in X_{-i}$, $m_{-i}^{X_{-i}}(s_{-i}) = m_{-i}^k(s_{-i}) \big/ \sum_{r_{-i} \in X_{-i}} m_{-i}^k(r_{-i})$, where $\text{supp}(m_{-i}^k) \cap X_{-i} \neq \emptyset$ and, $\forall \ell \in \{1,...,k\!-\!1\}$, $\text{supp}(m_{-i}^\ell) \cap X_{-i} = \emptyset$. Then $\{m_{-i}^{X_{-i}} \mid X_i \times X_{-i} \in H_i\}$ is a system of conjectures satisfying Bayes law. It remains to be shown that $\forall X_i \times X_{-i} \in H_i(p_i)$, $\forall s_i \in X_i$, $u_i(p_i, m_{-i}^{X_{-i}}(\cdot)) \geq u_i(s_i, m_{-i}^{X_{-i}}(\cdot))$. Suppose to the contrary that there exist $Y_i \times Y_{-i} \in H_i(p_i)$ and $r_i \in Y_i$ such that $u_i(p_i, m_{-i}^{Y_{-i}}(\cdot)) < u_i(r_i, m_{-i}^{Y_{-i}}(\cdot))$, where $\text{supp}(m_{-i}^k) \cap Y_{-i} \neq \emptyset$ and, $\forall \ell \in \{1,...,k\!-\!1\}$, $\text{supp}(m_{-i}^\ell) \cap Y_{-i} = \emptyset$. It follows from Mailath et al. (1993, Def. 2, Def. 3 & Thm. 1) that $Y_i \times Y_{-i}$ is a *strategic independence* for player i in the sense that, $\forall p_i, s_i \in Y_i$, $\exists r_i \in Y_i$ such that, $\forall s_{-i} \in Y_{-i}$, $u_i(r_i, s_{-i}) = u_i(s_i, s_{-i})$ and, $\forall s_{-i} \in S_{-i} \backslash Y_{-i}$, $u_i(r_i, s_{-i}) = u_i(p_i, s_{-i})$. Hence, r_i can be chosen such that $\forall s_{-i} \in S_{-i} \backslash Y_{-i}$, $u_i(r_i, s_{-i}) = u_i(p_i, s_{-i})$. By construction of $m_{-i}^{Y_{-i}}$ and r_i, either (a) both p_i and r_i are in Z_i^{k-1}, in which case it follows that $p_i \notin Z_i^k \supseteq Z_i^K$ (since $u_i(p_i, m_{-i}^{Y_{-i}}(\cdot)) < u_i(r_i, m_{-i}^{Y_{-i}}(\cdot))$ and, $\forall s_{-i} \in \text{supp}(m_{-i}^k) \backslash Y_{-i}$, $u_i(r_i, s_{-i}) = u_i(p_i, s_{-i})$ imply that $u_i(p_i, m_{-i}^k(\cdot)) < u_i(r_i, m_{-i}^k(\cdot))$, or (b) both p_i and r_i are not in Z_i^{k-1}, in which case $p_i \notin Z_i^{k-1} \supseteq Z_i^K$. •

Proofs of Propositions 2 and 3. Given the monotonicity of α_i^0, Propositions 2 and 3 are straightforward consequences of the states-of-the-world model of Section 2. They are therefore provided without proof. Proofs are available on request from the authors.

If $\emptyset \neq X_{-i} \subseteq S_{-i}$, let $\bar{a}_i(X_{-i}) := \{p_i \in S_i |\ \exists (m_{-i}^1, \ldots, m_{-i}^K) \in \mathbf{L}\Delta^0(S_{-i})$ with $m_{-i}^1 \in \Delta(X_{-i})$ such that p_i is an admissible best response to $(m_{-i}^1, \ldots, m_{-i}^K)\}$. If $\emptyset \neq X'_{-i} \subseteq X''_{-i} \subseteq S_{-i}$, then $\emptyset \neq \bar{a}_i(X'_{-i}) \subseteq \bar{a}_i(X''_{-i}) \subseteq S_i$. If $\emptyset \neq X = X_1 \times \ldots \times X_n \subseteq S$, write $\bar{a}(X) := \bar{a}_1(X_{-1}) \times \ldots \times \bar{a}_n(X_{-n})$. By Brandenburger (1992), p_i is a permissible strategy iff there exists $X = X_1 \times \ldots \times X_n$ with $p_i \in$

X_i such that $X \subseteq \bar{a}(X)$. If $P = P_1 \times ... \times P_n$ denotes the set of permissible strategy vectors, then $P = \bar{a}(P)$.

Proof of Prop. 5. Using Proposition 3(ii), Definitions 1 and 3 imply, $\forall i \in N$, $\bar{P}_i^0 := \bigcup_{\sigma_i \in \Pi_i^0} \sigma_i$ $= \bigcup_{\sigma_i \in \alpha_i^0(\Pi_{-i}^0)} \sigma_i \subseteq \bar{a}_i(\bar{P}_{-i}^0)$. Since $\bar{P}^0 \subseteq \bar{a}(\bar{P}^0)$ implies $\bar{P}^0 \subseteq P$, it follows that, $\forall i \in N$, $\bigcup_{\sigma_i \in \Pi_i^0} \sigma_i \subseteq P_i$. •

Proof of Prop. 6. Write $\Pi_i := \{\pi_i \in \Sigma_i | \ \pi_i$ is a permissible set for $i\}$. *(If)* Using the analogue to Proposition 3(ii), Definitions 1 and 2 imply, $\forall i \in N$, $\bar{P}_i := \bigcup_{\sigma_i \in \Pi_i} \sigma_i$ $=$ $\bigcup_{\sigma_i \in \alpha_i(\Pi_{-i})} \sigma_i \subseteq \bar{a}_i(\bar{P}_{-i})$. Since $\bar{P} \subseteq \bar{a}(\bar{P})$ implies $\bar{P} \subseteq P$, it follows that, $\forall i \in N$, $\bigcup_{\sigma_i \in \Pi_i} \sigma_i \subseteq P_i$. *(Only if)* $\bar{\Pi} := \{(P_1, \ldots, P_n)\}$ satisfies $\bar{\Pi} = \alpha(\bar{\Pi})$ since $P = \bar{a}(P)$. By the analogue to Proposition 2, $\forall i \in N$, P_i is a permissible set for i. Hence, $\forall i \in N$, $\bigcup_{\sigma_i \in \Pi_i} \sigma_i \supseteq P_i$. •

If $\emptyset \neq X_{-i} \subseteq S_{-i}$, let $\bar{b}_i(X_{-i}) := \{r_i \in S_i | \ \exists\, m_{-i} \in \Delta(X_{-i})$ such that r_i is a best response to $m_{-i}\}$. If $\emptyset \neq X'_{-i} \subseteq X''_{-i} \subseteq S_{-i}$, then $\emptyset \neq \bar{b}_i(X'_{-i}) \subseteq \bar{b}_i(X''_{-i}) \subseteq S_i$. If $\emptyset \neq X = X_1 \times ... \times X_n \subseteq S$, write $\bar{b}(X) := \bar{b}_1(X_{-1}) \times ... \times \bar{b}_n(X_{-n})$. By Bernheim (1984) and Pearce (1984) (but note that we here consider pure strategies only and allow conjectures to be correlated), r_i is a rationalizable strategy iff there exists $X = X_1 \times ... \times X_n$ with $r_i \in X_i$ such that $X \subseteq \bar{b}(X)$. Pearce (1984) says that X satisfies the *best response property* if $X \subseteq \bar{b}(X)$. If $R = R_1 \times ... \times R_n$ denotes the set of rationalizable strategy vectors, then $R = \bar{b}(R)$.

Proof of Prop. 4. Using Proposition 3(ii), Definitions 1, 3 and 6 imply, $\forall i \in N$, $\bar{P}_i^0 :=$ $\bigcup_{\sigma_i \in \Pi_i^0} \sigma_i = \bigcup_{\sigma_i \in \alpha_i^0(\Pi_{-i}^0)} \sigma_i \subseteq \bar{b}_i(\bar{P}_{-i}^0)$. Since $\bar{P}^0 \subseteq \bar{b}(\bar{P}^0)$ implies $\bar{P}^0 \subseteq R$, it follows that, $\forall i \in N$, $\bigcup_{\sigma_i \in \Pi_i^0} \sigma_i \subseteq R_i$. •

Proof of Prop. 7. Write $\mathrm{P}_i := \{\rho_i \in \Sigma_i | \ \rho_i$ is a rationalizable set for $i\}$. *(If)* Using the analogue to Proposition 3(ii), Definitions 2 and 6 imply, $\forall i \in N$, $\bar{R}_i := \bigcup_{\sigma_i \in \mathrm{P}_i} \sigma_i = \bigcup_{\sigma_i \in \beta_i(\mathrm{P}_{-i})} \sigma_i \subseteq$ $\bar{b}_i(\bar{R}_{-i})$. Since $\bar{R} \subseteq \bar{b}(\bar{R})$ implies $\bar{R} \subseteq R$, it follows that, $\forall i \in N$, $\bigcup_{\sigma_i \in \mathrm{P}_i} \sigma_i \subseteq R_i$. *(Only if)* $\bar{\mathrm{P}} := \{(R_1, \ldots, R_n)\}$ satisfies $\bar{\mathrm{P}} = \beta(\bar{\mathrm{P}})$ since $R = \bar{b}(R)$. By the analogue to Proposition 2, $\forall i \in N$, R_i is a rationalizable set for i. Hence, $\forall i \in N$, $\bigcup_{\sigma_i \in \mathrm{P}_i} \sigma_i \supseteq R_i$. •

APPENDIX B: DERIVATIONS FOR THE EXAMPLES

The algorithm of Proposition 3(iii) is used to determine the collection of vectors of fully permissible sets in $G_1 - G_6$.

G_1:

$\Xi(0) = \Sigma = \Sigma_1 \times \Sigma_2$

$\Xi(1) = \{\{U\},\{M\},\{U,M\}\} \times \Sigma_2$

$\Xi(2) = \{\{U\},\{M\},\{U,M\}\} \times \{\{L\},\{L,R\}\}$

$\Pi^0 = \Xi(3) = \{\{U\},\{U,M\}\} \times \{\{L\},\{L,R\}\}$

G_2:

$\Xi(0) = \Sigma = \Sigma_1 \times \Sigma_2$

$\Xi(1) = \{\{U,M\}\} \times \Sigma_2$

$\Pi^0 = \Xi(2) = \{\{U,M\}\} \times \{\{L\}\}$

G_3:

$\Xi(0) = \Sigma = \Sigma_1 \times \Sigma_2$

$\Xi(1) = \{\{U\},\{M\},\{U,M\}\} \times \Sigma_2$

$\Xi(2) = \{\{U\},\{M\},\{U,M\}\} \times \{\{L\},\{L,R\}\}$

$\Xi(3) = \{\{M\},\{U,M\}\} \times \{\{L\},\{L,R\}\}$

$\Xi(4) = \{\{M\},\{U,M\}\} \times \{\{L\}\}$

$\Pi^0 = \Xi(5) = \{\{M\}\} \times \{\{L\}\}$

G_4:

$\Xi(0) = \Sigma = \Sigma_1 \times \Sigma_2$

$\Xi(1) = \{\{NU\},\{ND\},\{BU\},\{NU,ND\},\{ND,BU\},\{NU,BU\},\{NU,ND,BU\}\} \times \Sigma_2$

$\Xi(2) = \{\{NU\},\{ND\},\{BU\},\{NU,ND\},\{ND,BU\},\{NU,BU\},\{NU,ND,BU\}\} \times$

$\{\{LL\},\{RL\},\{LL,LR\},\{RL,RR\},\{LL,RL\},\{LL,LR,RL,RR\}\}$

$\Xi(3) = \{\{NU\},\{BU\},\{ND,BU\},\{NU,BU\},\{NU,ND,BU\}\} \times$

$\{\{LL\},\{RL\},\{LL,LR\},\{RL,RR\},\{LL,RL\},\{LL,LR,RL,RR\}\}$

$\Xi(4) = \{\{NU\},\{BU\},\{ND,BU\},\{NU,BU\},\{NU,ND,BU\}\} \times$

$\{\{LL\},\{RL\},\{LL,LR\},\{LL,RL\}\}$

$\Xi(5) = \{\{NU\},\{BU\},\{NU,BU\}\} \times \{\{LL\},\{RL\},\{LL,LR\},\{LL,RL\}\}$

$\Xi(6) = \{\{NU\},\{BU\},\{NU,BU\}\} \times \{\{LL\},\{LL,LR\},\{LL,RL\}\}$

$\Xi(7) = \{\{NU\},\{NU,BU\}\} \times \{\{LL\},\{LL,LR\},\{LL,RL\}\}$

$\Xi(8) = \{\{NU\},\{NU,BU\}\} \times \{\{LL\},\{LL,LR\}\}$

$\Xi(9) = \{\{NU\}\} \times \{\{LL\},\{LL,LR\}\}$

$\Pi^0 = \Xi(10) = \{\{NU\}\} \times \{\{LL,LR\}\}$

G_5:

$\Xi(0) = \Sigma = \quad \Sigma_1 \times \Sigma_2$

$\Xi(1) = \{\{M\},\{D\},\{M,D\}\} \times \Sigma_2$

$\Xi(2) = \{\{M\},\{D\},\{M,D\}\} \times \{\{L\},\{L,R\}\}$

$\Pi^0 = \Xi(3) = \{\{D\},\{M,D\}\} \times \{\{L\},\{L,R\}\}$

G_6:

$\Xi(0) = \Sigma = \quad \Sigma_1 \times \Sigma_2$

$\Xi(1) = \{\{U\}\} \times \Sigma_2$

$\Pi^0 = \Xi(2) = \{\{U\}\} \times \{\{L\}\}$

REFERENCES

Asheim, G.B., 1994, "Defining Rationalizability in 2–Player Extensive Games", Memorandum No. 25, Department of Economics, University of Oslo.

Aumann, R., 1995, "Backward Induction and Common Knowledge of Rationality", *Games and Economic Behavior* 8, 6–19.

Aumann, R. and A. Brandenburger, 1995, "Epistemic Conditions for Nash Equilibrium", *Econometrica* 63, 1161–1180.

Basu, K., 1990, "On the Non-Existence of a Rationality Definition for Extensive Games", *International Journal of Game Theory* 19, 33–44.

Battigalli, P., 1989, "Algorithmic Solutions for Extensive Games", in Ricci (ed.) *Decision Processes in Economics Proceedings*, Springer.

Battigalli, P., 1994, "Strategic Rationality Orderings and the Best Rationalization Principle", forthcoming in *Games and Economic Behavior*.

Battigalli, P., 1995, "On Rationalizability in Extensive Games", mimeo, Princeton University.

Ben-Porath, E., 1994, "Rationality, Nash Equilibrium, and Backwards Induction in Perfect Information Games", mimeo, Northwestern University.

Ben-Porath, E. and E. Dekel, 1992, "Coordination and the Potential for Self-Sacrifice", *Journal of Economic Theory* 57, 36–51.

Bernheim, D., 1984, "Rationalizable Strategic Behavior", *Econometrica* 52, 1007–1028.

Bicchieri, C., 1989, "Self-Refuting Theories of Strategic Interaction: A Paradox of Common Knowledge", *Erkenntnis* 30, 69–85.

Binmore, K., 1987, "Modelling Rational Players I", *Economics and Philosophy* 3, 179–214.

Binmore, K., 1995, "Backward Induction and Rationality", Discussion Paper 95–10, University College London.

Blume, L., A. Brandenburger, and E. Dekel, 1991, "Lexicographic Probabilities and Choice under Uncertainty", *Econometrica* 59, 61–79.

Börgers, T., 1991, "On the Definition of Rationalizability in Extensive Form Games", Discussion Paper 91–22, University College London.

Börgers, T., 1994, "Weak Dominance and Approximate Common Knowledge", *Journal of Economic Theory* 64, 265–276.

Börgers, T. and L. Samuelson, 1992, "'Cautious' Utility Maximization and Iterated Weak Dominance", *International Journal of Game Theory* 21, 13–25.

Brandenburger, A., 1992, "Lexicographic Probabilities and Iterated Admissibility", in Dasgupta, Gale, Hart, Maskin (eds.), *Economic Analysis of Markets and Games*, MIT Press.

Dekel, E. and D. Fudenberg, 1990, "Rational Behavior with Payoff Uncertainty", *Journal of Economic Theory* 52, 243–267.

Dufwenberg, M., 1994, "Tie-Break Rationality and Tie-Break Rationalizability", Working Paper 1994:29, Dept. of Economics, Uppsala University.

Gul, F., 1995, "Rationality and Coherent Theories of Strategic Behavior", forthcoming in *Journal of Economic Theory*.

Hendon, E., H.J. Jacobsen, B. Sloth, and T. Tranæs, 1994, "Expected Utility with Lower Probabilities", *Journal of Risk and Uncertainty* 8, 197-216.

Kohlberg, E. and J.-F. Mertens, 1986, "On the Strategic Stability of Equilibria", *Econometrica* 54, 1003–1037.

Kreps, D. and R. Wilson, 1982, "Sequential Equilibria", *Econometrica* 50, 863–894.

Luce, D. and H. Raiffa, 1957, *Games and Decisions*, Wiley.

Mailath, G., L. Samuelson, and J. Swinkels, 1993, "Extensive Form Reasoning in Normal Form Games", *Econometrica* 61, 273–302.

Osborne, M.J. and A. Rubinstein, 1994, *A Course in Game Theory*, MIT Press.

Pearce, D., 1984, "Rationalizable Strategic Behavior and the Problem of Perfection", *Econometrica* 52, 1029–1050.

Rajan, U., 1993, "Non-Archimedean Probabilities, Equilibrium Refinements, and Rationalizability", mimeo, Stanford University.

Reny, P., 1993, "Common Belief and the Theory of Games with Perfect Information", *Journal of Economic Theory* 59, 257–274.

Rubinstein, A., 1991, "Comments on the Interpretation of Game Theory", *Econometrica* 59, 909–924.

Samuelson, L., 1992, "Dominated Strategies and Common Knowledge", *Games and Economic Behavior* 4, 284–313.

Stahl, D., 1993, "Lexicographic Rationality, Common Knowledge, and Iterated Admissibility", mimeo, University of Texas.

Tan, T. and S.R.C. Werlang, 1988, "The Bayesian Foundations of Solution Concepts of Games", *Journal of Economic Theory* 45, 370–391.

van Damme, E., 1989, "Stable Equilibria and Forward Induction", *Journal of Economic Theory* 48, 476–496.

Veronesi, P., 1995, "Lexicographic Rationality Orderings and Iterative Weak Dominance", mimeo, Harvard University.