ECONSTOR

Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Asheim, Geir B.; Dufwenberg, Martin

Working Paper

Rational Reasoning and Rationalizable Sets

Discussion Paper, No. 1129

Provided in Cooperation with:
Kellogg School of Management - Center for Mathematical Studies in Economics and Management Science, Northwestern University

Suggested Citation: Asheim, Geir B.; Dufwenberg, Martin (1995) : Rational Reasoning and Rationalizable Sets, Discussion Paper, No. 1129, Northwestern University, Kellogg School of Management, Center for Mathematical Studies in Economics and Management Science, Evanston, IL

This Version is available at:
http://hdl.handle.net/10419/221485

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Discussion Paper No. 1129

Rational Reasoning and Rationalizable Sets*

by Geir B. Asheim** and Martin Dufwenberg***

June 1995

Abstract

Earlier contributions have shown that imposing common knowledge of rationality is problematic when rationality is defined as choosing an admissible best response. Here we instead impose common knowledge of rational *reasoning* and define the concept of *rationalizable sets*. General existence (for any non-empty valued best response operator) is established, and a finite algorithm (eliminating strategy sets instead of strategies) is provided. Combined with the ordinary best response operator, Bernheim-Pearce rationalizability is fully characterized. Combined with the admissible best response operator, rationalizability is defined under the assumption of cautious and sequentially rational behavior, and a notion of forward induction is captured.

---

* We thank Elchanan Ben-Porath, Eddie Dekel, Faruk Gul, Phil Reny, Larry Samuelson, Jeroen Swinkels, and Jörgen Weibull for valuable discussions and comments. Asheim gratefully acknowledges the hospitality of Stockholm, Munich, and Northwestern Universities and financial support from the Norwegian Research Council. Finally, the acknowledgments in Asheim (1994) apply also here; in particular, thanks go to Yossi Greenberg, Ariel Rubinstein, and Benyamin Shitovitz for helpful discussions and McGill University for hospitality.

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

*** Department of Economics, Uppsala University, Box 513, S-75120 Uppsala, Sweden (Tel. +46-18-181593, Fax. +46-18-181590/181478, Internet: martin.dufwenberg@nek.uu.se).

1. INTRODUCTION

Bernheim (1984, Definitions 3.2 & 3.3) and Pearce (1984, Definitions 1 & 3) define rationalizability in a normal form game by imposing that it be common knowledge that each player chooses rationally. They define a choice to be rational if it is an ordinary best response, implying that the choice of a weakly dominated strategy is not necessarily irrational. It would seem desirable instead to define a choice to be rational if it is an admissible best response since this would exclude such incautious behavior from being rational.[1] However, the project of combining common knowledge of rational choice with admissibility easily runs into problems. Samuelson (1992) presents a thorough investigation of these problems and concludes that, for some games, common knowledge of rational choice is inconsistent with admissibility. The following example is a simple illustration of this inconsistency.

EXAMPLE 1: Samuelson (1992, Ex. 8; also included as Ex. 2 in Börgers & Samuelson, 1992).

	L	***R***	
U	1,1	1,0	
D	1,0	0,1	G_1

Only U is an admissible best response for 1. Hence, if 2 chooses an admissible best response and knows that 1 chooses an admissible best response, then 2 knows that 1 chooses U and only L is an admissible best response for 2. Now, if 1 chooses an admissible best response and knows that 2 chooses an admissible best response and knows that 2 knows that 1 chooses an admissible best response, then 1 knows that 2 chooses L. However, if 1 *knows* that 2 does not choose R, then both U and D appear to be admissible best responses for 1. The problem is: Why should 1 be cautious if 1 knows that 2 knows that 1 is cautious?[2]

---

[1] The prescription that players should avoid playing weakly dominated strategies is argued by e.g. Luce & Raiffa (1957), Kohlberg & Mertens (1986), and Dekel & Fudenberg (1990).

[2] Cubitt & Sugden (1994) present a similar example and make a similar argument.

Our suggestion for resolving the inconsistency of common knowledge and admissibility is to change the object for the common knowledge: Instead of imposing common knowledge of rational choice, we impose common knowledge of rational *reasoning*. We will seek to show that by modeling the reasoning of the players rather than their choice, no logical problems are encountered when common knowledge is combined with admissibility.

To motivate our modeling of the players' reasoning, note that if a player makes prescriptive use of Bernheim's (1984) Definition 3.2, then his reasoning will determine a subset of his set of feasible conjectures. This set consists of independent conjectures with support included in the set of the opponents' rationalizable pure strategy profiles. Here we accommodate alternative best response operators by assuming that each player has not only a primary conjecture, but also a secondary conjecture in the hypothetical case that his opponents make a choice inconsistent with the primary conjecture, etc. Such a hierarchy of non-overlapping conjectures can be derived from Myerson's (1986) concept of a *conditional probability system*. Select some best response operator. In analogy with the prescriptive use of Bernheim's (1984) Definition 3.2, assume that the reasoning of each player leads to *some* subset of his set of conditional probability systems, being the set of conditional probability systems that are consistent with his reasoning about the strategic situation. This is turn determines the player's *best response set*, defined as the set of strategies that each is a best response to some conditional probability system in this subset.

A player realizes that the reasoning of each opponent will determine a best response set. Say that the player *reasons rationally* if he finds it infinitely more likely that each opponent chooses a strategy in her best response set rather than a strategy outside her best response set. Since a player may be uncertain as to *which* set is an opponent's actual best response set, we assume that the player has an independent probability distribution over vectors of sets that can possibly be the vectors of his opponents' actual best response sets. Furthermore, let rational reasoning be appropriately generalized to yield a set of conditional probability systems that are consistent with this probability distribution. A *rationalizable set* is a non-empty set of strategies that can be a best response set if it is common knowledge that all players reason rationally.

The concepts of rational reasoning and rationalizable sets can be illustrated in Example 1 for the case of the admissible best response operator. The only possible best response sets for player 2 are $\{L\}$, $\{R\}$, and $\{L,R\}$. If 1 assigns probability 1 to $\{L\}$ being 2's best response set, then 1 finds it infinitely more likely that 2 chooses L rather than R, thereby determining a unique conditional probability system. Since an admissible best response to a conditional probability system satisfies lexicographic optimization (see Blume et al. (1991) and Definition 4 of Section 5), U is the unique admissible best response to this conditional probability system. For any other probability distribution over 2's possible best response sets, U is clearly the unique admissible best response to any conditional probability system consistent with the probability distribution. Hence, if 1 reasons rationally, $\{U\}$ is his best response set. If 2 reasons rationally and knows that 1 reasons rationally, then 2 knows that $\{U\}$ is 1's best response set, leading her[3] to find it infinitely more likely that 1 plays U rather than D. This determines a unique conditional probability system with L being the unique admissible best response to this conditional probability system. Hence, if 2 reasons rationally and knows that 1 reasons rationally, then $\{L\}$ is her best response set. If 1 reasons rationally and knows that 2 reasons rationally and knows that 2 knows that 1 reasons rationally, then 1 knows that $\{L\}$ is 2's best response set, leading him to find it infinitely more likely that 2 plays L rather than R. As argued above, $\{U\}$ is still 1's best response set.

Hence, while imposing common knowledge of rational choice leads to iterated elimination of *strategies*, the above procedure illustrates how imposing common knowledge of rational reasoning leads to iterated elimination of strategy *sets*. In particular, since the *strategy* R is *not* eliminated (rather, the procedure eliminates the strategy *sets* $\{R\}$ and $\{L,R\}$), common knowledge of rational reasoning does not prevent player 1 from taking into account the possibility that 2 will play R. With the admissible best response operator, sets that survive the iterated elimination will be referred to as *admissible* rationalizable sets. Thus, in Example 1 $\{U\}$ is the unique admissible rationalizable set for player 1 and $\{L\}$ is the unique admissible rationalizable set for player 2.

---

[3] In two-player games, we refer to 1 as the male player and 2 as the female player.

The "Burning Money" game (included as Example 5 of Section 6) due to Ben-Porath & Dekel (1992, Fig. 1.2) and van Damme (1989, Fig. 5) illustrates how the elimination of strategy sets rather than strategies may have attractive consequences. When "Burning Money" is solved by iterated elimination of weakly dominated strategies, the first elimination by 2 requires her to interpret burning as an explicable action. Still, any strategy involving burning is eventually eliminated; hence, burning does not emerge as an explicable action. Thereby the validity of 2's interpretation is undermined. When this game is solved by iterated elimination of strategy sets, 2 need at no stage interpret burning as an explicable action. The reason is that 2 always has the option of assigning probability 1 to a strategy set for 1 in which no strategy involves burning. In fact, this set is the only set for 1 that survives the iterated elimination of strategy sets; hence, it is 1's unique admissible rationalizable set. As for iterated elimination of weakly dominated strategies, common knowledge of rational reasoning yields the forward induction *outcome*. Still, 2 does not interpret burning as an explicable outcome, and her unique admissible rationalizable set imposes no restriction on her action conditional on 1 burning.

In the above examples, the iterated elimination of strategy sets implied by the common knowledge of rational reasoning leads to a unique admissible rationalizable set for each player. However, in certain games (see e.g. Examples 3 and 4 of Section 6 as well as Reny's (1993) "Take-it-or-leave-it") where there happens to be no game-theoretic consensus on a unique set-valued prescription, it turns out that there are multiple admissible rationalizable sets for each player. Thus, in the present framework, a player endogenizes the prescriptive ambiguity by being uncertain as to which admissible rationalizable set is the actual best response set of an opponent.

For a fixed non-empty valued best response operator, Section 3 formalizes common knowledge of rational reasoning, and defines, characterizes and provides a finite algorithm for the concept of rationalizable set. This algorithm does not eliminate strategies; rather, it eliminates strategy *sets* that cannot be best response sets. It is established that, for each player, there exists at least one rationalizable set. If the ordinary best response operator is applied, Section 4 establishes that a strategy is contained in some ordinary rationalizable set iff it is a rationalizable

pure strategy as defined by Bernheim (1984) and Pearce (1984). Hence, common knowledge of rational reasoning combined with the ordinary best response operator fully characterizes Bernheim-Pearce rationalizability. In Section 5, the admissible best response operator is introduced. Using Mailath et al.'s (1993) concept of a *strategic independence*, it is shown how admissibility implies sequential rationality in any underlying extensive form. It is established that combining common knowledge of rational reasoning with the admissible best response operator refines Bernheim-Pearce rationalizability: Any admissible rationalizable set is included in the set of rationalizable pure strategies as defined by Bernheim (1984) and Pearce (1984). In Sections 6 and 7 we consider five examples to illustrate the properties and applicability of the concept of admissible rationalizable sets, while Section 8 concludes. All proofs are relegated to Appendix A, while Appendix B contains derivations for the examples.

In closing this introduction, we would like to stress that this application of admissibility deals directly with two types of imperfect behavior discussed by Pearce (1984): implausible behavior at unreached information sets and incautious optimization. To cater for the first imperfection Pearce develops his solution concept 'extensive form rationalizability' in which players exploit information embodied in extensive form information sets. To cater for the second he requires players' conjectures to have full support. In contrast, the present framework of rational reasoning leads to an admissible best response being defined by lexicographic optimization. This implies optimization at all extensive form information sets, and it entails cautiousness without constraining players to full support conjectures.

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. As it turns out, once we insist on lexicographic optimization, sequential rationality is nevertheless adequately captured. For this reason, the results we report have a bearing on any given extensive game although our formal definitions deal with games represented in normal form.

2. PRELIMINARIES

With $N = \{1,...,n\}$ as the set of players, let S_i^* denote player i's finite set of pure strategies, and let $u_i^*: S^* \to \Re$ be i's payoff function, where $S^* \equiv S_1^* \times ... \times S_n^*$. Then $G^* = (S^*, u^*)$ is a normal form game. Let $G = (S,u)$ be the corresponding pure strategy reduced normal form (PRNF) game (Mailath et al., 1993, Def. 1), with $S \equiv S_1 \times ... \times S_n \equiv S_i \times S_{-i}$, where $-i$ denotes $N\backslash\{i\}$. Throughout, $\subseteq$ ($\subset$) denotes weak (strict) set inclusion. If $\varnothing \neq X_{-i} \subseteq S_{-i}$, 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 finite sets of "states" other than S_{-i}). Let $p_{-i}(s_{-i}|X_{-i})$ be the (subjective) probability assigned by i to s_{-i} conditional on X_{-i}; i.e., $p_{-i}(s_{-i}|X_{-i}) \in \Delta(X_{-i})$. Abuse notation slightly by writing $u_i(s_i, p_{-i}(\cdot|X_{-i}))$ for i's expected payoff given the (subjective) probability distribution $p_{-i}(\cdot|X_{-i})$. Say that $p_{-i}(\cdot|\cdot)$ is a *conditional probability system* if there exists a sequence of probability distributions $\{\hat{p}_{-i}^k\}_{k=1}^{\infty}$ in $\Delta^0(S_{-i})$ such that $\forall(\varnothing \neq) X_{-i} \subseteq S_{-i}$ and $\forall s_{-i} \in S_{-i}$, $p_{-i}(s_{-i}|X_{-i}) = \lim_{k\to\infty} \hat{p}_{-i}^k(s_{-i}) \big/ \sum_{t_{-i} \in X_{-i}} \hat{p}_{-i}^k(t_{-i})$ if $s_{-i} \in X_{-i}$ and $p_{-i}(s_{-i}|X_{-i}) = 0$ if $s_{-i} \in S_{-i} \backslash X_{-i}$. Say that $p_{-i}(\cdot|\cdot)$ is an *independent* conditional probability system if, in addition, $\forall j \neq i$, $\exists \{\hat{p}_j^k\}_{k=1}^{\infty}$ in $\Delta^0(S_j)$ such that, $\forall k \geq 1$, $\hat{p}_{-i}^k(s_{-i}) \equiv \prod_{j\neq i} \hat{p}_j^k(s_j)$ if $s_{-i} \equiv (s_j)_{j\neq i}$.

Remark: By Myerson (1986, Theorem 1), taking the limit of full support conjectures is sufficient and necessary for a conditional probability system satisfying Bayes' law. In the present context, this use of full support conjectures should not be given any behavioral interpretation; e.g., that players make mistakes. Players making mistakes is *not* part of the subsequent analysis since (a) players only optimize given the conditional probability system obtained in the limit, and (b) in the limit, there are no full support restrictions on the conditional probability systems. The independence condition above is strong; see Kohlberg & Reny (1992) and Swinkels (1994) for defenses of its appropriateness. Independence is assumed since it implies that if a player observes that either one or two of his opponents have made a choice to which he assigns probability 0, then he finds it infinitely more likely that only one has done so (rather than both). The assumption is dispensable.

3. COMMON KNOWLEDGE OF RATIONAL REASONING

The purpose of the present section is to construct a framwork for analyzing the reasoning of the players. The analysis presupposes that a best response operator has been selected. The precise nature of the best response operator is not specified here. The only requirement on the best response operator imposed by the analysis is that, for each player i, if $p_{-i}(\cdot|\cdot)$ is a conditional probability system, then there exists a best response to $p_{-i}(\cdot|\cdot)$. In the subsequent sections, the implications of the analysis are explored using various best response operators.

The modeling of the reasoning of each player i will provide guidelines for determining the *best response set* for i, being the set of strategies from which i may choose given that he chooses a best response to some conditional probability system consistent with his reasoning about the strategic situation. Player i realizes that the reasoning of each opponent j will likewise lead to a best response set for j. Player i finds it infinitely more likely that each opponent j will choose a strategy in rather than outside her best response set.

Allow i to be uncertain as to which subset of S_j is j's actual best response set. Writing $\Sigma_j := 2^{S_j} \setminus \{\varnothing\}$, let $\pi_j(\cdot) \in \Delta(\Sigma_j)$ be a probability distribution having the interpretation that i assigns probability $\pi_j(\sigma_j)$ to σ_j being j's actual best response set. Note that $\pi_j(\cdot)$ expresses player i's uncertainty concerning the reasoning—not the choice—of opponent j. Impose that, $\forall \sigma_j', \sigma_j'' \in \Sigma_j$, i finds it infinitely more likely that j chooses $r_j \in \sigma_j'$ conditional on σ_j' being j's best response set, rather than j chooses $s_j \in S_j \setminus \sigma_j''$ conditional on σ_j'' being j's best response set. Writing $\Sigma \equiv \Sigma_1 \times \ldots \times \Sigma_n \equiv \Sigma_i \times \Sigma_{-i}$, this motivates the following definition.

DEFINITION 1. Say that $p_i(\cdot|\cdot)$ is *consistent* with $\pi_{-i}(\cdot) \equiv \prod_{j \neq i} \pi_j(\cdot) \in \Delta(\Sigma_{-i})$ if $p_i(\cdot|\cdot)$ is an independent conditional probability system that is generated by $\{\prod_{j \neq i} \hat{p}_j^k\}_{k=1}^{\infty}$ satisfying, $\forall j \neq i$, $\forall \sigma_j \in \Sigma_j$, $\exists \{\tilde{p}_j^k(\sigma_j)\}_{k=1}^{\infty}$ in $\Delta^{\#}(S_j)$ such that

1. $r_j \in \sigma_j'$ and $s_j \in S_j \setminus \sigma_j''$ imply $\lim_{k \to \infty} \tilde{p}_j^k(\sigma_j'')(s_j) / \tilde{p}_j^k(\sigma_j')(r_j) = 0$,
2. $\forall k \geq 1$, $\forall s_j \in S_j$, $\hat{p}_j^k(s_j) = \sum_{\hat{\sigma}_j \in \Sigma_j} \pi_j(\hat{\sigma}_j)\, \tilde{p}_j^k(\hat{\sigma}_j)(s_j)$.

Note that the consistency of a conditional probability system for i depends only on best response sets for his opponents to which i assigns positive probability. Player i believes with probability 1 that each opponent j will choose a strategy in the union of the best response sets to which i assigns positive probability. Conditional on j choosing outside the union of the best response sets to which i assigns positive probability, Definition 1 yields no constraint on conjectures, reflecting that—according to i's reasoning—j is making an inexplicable choice.

If $\pi_{-i}(\cdot)$ is an independent probability distribution in $\Delta(\Sigma_{-i})$, then $h_i(\pi_{-i}(\cdot)) := \{r_i \in S_i | \exists\ p_{-i}(\cdot|\cdot)$ consistent with $\pi_{-i}(\cdot)$ such that r_i is a best response to $p_{-i}(\cdot|\cdot)\}$ denotes i's best response set. For any independent probability distribution $\pi_{-i}(\cdot)$ in $\Delta(\Sigma_{-i})$, there exists $p_{-i}(\cdot|\cdot)$ consistent with $\pi_{-i}(\cdot)$; this is seen by letting, $\forall j \neq i$, $\forall \sigma_j \in \Sigma_j$, $\lim_{k\to\infty} \tilde{p}_j^k(\sigma_j)(s_j) > 0$ iff $s_j \in \sigma_j$. Furthermore, by assumption, there exists a best response to any $p_{-i}(\cdot|\cdot)$. Hence, $h_i(\pi_{-i}(\cdot)) \neq \varnothing$ for any independent probability distribution $\pi_{-i}(\cdot)$ in $\Delta(\Sigma_{-i})$. If P_{-i} is a nonempty rectangular subcollection of Σ_{-i}, let $\beta_i(\mathrm{P}_{-i}) := \{h_i(\pi_{-i}(\cdot)) | \pi_{-i}(\cdot) \equiv \prod_{j\neq i} \pi_j(\cdot) \in \Delta(\mathrm{P}_{-i})\}$ denote the collection that contains a strategy set for i iff it is a best response set to some independent probability distribution in $\Delta(\mathrm{P}_{-i})$. Note that $\beta_i(\cdot)$ is defined neither for the empty collection of sets (i.e., we require $\mathrm{P}_{-i} \neq \varnothing$) nor for a collection of sets that contains the empty set (i.e., we require $\varnothing \notin \mathrm{P}_{-i}$ since $\varnothing \notin \Sigma_{-i} \supseteq \mathrm{P}_{-i}$).

PROPOSITION 1. *$\forall i \in N$, (i) if $\varnothing \neq \mathrm{P}_{-i} \equiv \prod_{j\neq i} \mathrm{P}_j \subseteq \Sigma_{-i}$, then $\beta_i(\mathrm{P}_{-i}) \neq \varnothing$ and $\varnothing \notin \beta_i(\mathrm{P}_{-i})$, and (ii) if $\varnothing \neq \mathrm{P}'_{-i} \equiv \prod_{j\neq i} \mathrm{P}'_j \subseteq \mathrm{P}''_{-i} \equiv \prod_{j\neq i} \mathrm{P}''_j \subseteq \Sigma_{-i}$, then $\beta_i(\mathrm{P}'_{-i}) \subseteq \beta_i(\mathrm{P}''_{-i})$.*

Consider a nonempty rectangular subcollection of vectors of nonempty strategy sets, P, and write $\beta(\mathrm{P}) \equiv \beta_1(\mathrm{P}_{-1}) \times \ldots \times \beta_n(\mathrm{P}_{-n})$, where $\beta(\mathrm{P})$ is a nonempty rectangular subcollection of vectors of nonempty strategy sets by Proposition 1(i). For later reference, let $\beta^0(\mathrm{P}) := \mathrm{P}$, and, for $\forall k \geq 1$, define $\beta^k(\mathrm{P})$ inductively by $\beta^k(\mathrm{P}) := \beta(\beta^{k-1}(\mathrm{P}))$. In analogy with the terminology of Greenberg (1990), say that P is *internally stable* if $\mathrm{P} \subseteq \beta(\mathrm{P})$, *externally stable* if $\mathrm{P} \supseteq \beta(\mathrm{P})$, and *stable* if $\mathrm{P} = \beta(\mathrm{P})$. Internal stability means that any set in P_i is a best response set given some independent probability distribution in $\Delta(\mathrm{P}_{-i})$. External stability means that any best response set given some independent probability distribution in $\Delta(\mathrm{P}_{-i})$ is a set in P_i.

To define the concept of *rational reasoning* and to explore the implications of common knowledge of rational reasoning, consider the following epistemological analysis. A *state* ω is a complete description of the reasoning of the players, including how the players reason about the opponents' reasoning, how the players reason that the opponents reason about their opponents' reasoning, etc. However, a state does not determine the actual strategy choices of the players. This seems appropriate as the present analysis models each player's reasoning, not his choice. Each player is assumed to reason independently of the game actually being played. Thus, what this reasoning yields does not depend on the actual outcome of the game if it were to be played.

Formally, with Ω denoting the *state space*, each state $\omega\in\Omega$ specifies for each player i

- $\Pi_i(\omega) \subseteq \Omega$, which denotes i's set of possible states given ω. Here, $\Pi_i(\cdot)$ is *partitional* in the sense that there is a partition of Ω such that, $\forall\omega\in\Omega$, $\Pi_i(\omega)$ is the element of the partition that contains ω. Say that i *knows* the event $E \subseteq \Omega$ given ω if $\Pi_i(\omega) \subseteq E$. Since $\omega\in\Pi_i(\omega)$, it follows that an event is true ($\omega\in E$) if i knows it.
- $\rho_i(\omega) \in \Sigma_i$, which denotes i's best response set given ω. This is the set of strategies from which i may choose given that he chooses a best response to some conditional probability system consistent with his reasoning about the strategic situation. Since $\forall\omega'\in\Pi_i(\omega)$ are indistinguishable for i, we have that, $\forall\omega'\in\Pi_i(\omega)$, $\rho_i(\omega') = \rho_i(\omega)$.

Since each player i thinks that his opponents reason independently of each other, we require that, $\forall\omega \in \Omega$, $P'_{-i}(\omega) := \{\rho_{-i}(\omega')|\ \omega'\in\Pi_i(\omega)\}$ is rectangular. Hence, $\forall\omega \in \Omega$, $\rho_{-i}(\omega) \in P'_{-i}(\omega) \equiv \prod_{j\neq i} P'_j(\omega) \subseteq \Sigma_{-i}$ since $\omega \in \Pi_i(\omega)$, and $P'_{-i}(\omega') = P'_{-i}(\omega)$ if $\omega' \in \Pi_i(\omega)$ since $\omega' \in \Pi_i(\omega)$ implies $\Pi_i(\omega') = \Pi_i(\omega)$. The interpretation of $P'_{-i}(\omega) \equiv \prod_{j\neq i} P'_j(\omega)$ is that i knows given ω that, for each opponent j, j's actual best response set is in $P'_j(\omega)$.

Say that the event $E \subseteq \Omega$ is *mutual knowledge* given ω if, $\forall i\in N$, $\Pi_i(\omega) \subseteq E$. Write $\Pi(\omega) := \bigcup_{i\in N} \Pi_i(\omega)$. Then the event $E \subseteq \Omega$ is mutual knowledge given ω iff $\Pi(\omega) \subseteq E$. If $\Phi \subseteq \Omega$, write $\Pi^0(\Phi) := \Phi$ and, $\forall k\geq 1$, $\Pi^k(\Phi) := \bigcup_{\omega\in\Pi^{k-1}(\Phi)} \Pi(\omega)$. Say that the event $E \subseteq \Omega$ is *common knowledge* given ω if, $\forall k\geq 0$, $\bigcup_{m=0}^{k} \Pi^k(\{\omega\}) \subseteq E$. Then, since, $\forall k\geq 1$, $\Pi^k(\Phi) \supseteq \Pi^{k-1}(\Phi)$, the event $E \subseteq \Omega$ is common knowledge given ω iff $\lim_{k\to\infty} \Pi^k(\{\omega\}) \subseteq E$.

Say that *i reasons rationally* given ω if $\rho_i(\omega) \in \beta_i(\mathrm{P}'_i(\omega))$. Let $\mathrm{E}_i^* := \{\omega \in \Omega |$ *i* reasons rationally given $\omega\}$, and let $\mathrm{E}^* := \bigcap_{i \in N} \mathrm{E}_i^*$. Then it is common knowledge given ω that, $\forall i \in N$, *i* reasons rationally iff $\lim_{k \to \infty} \Pi^k(\{\omega\}) \subseteq \mathrm{E}^*$. Imposing that *i* reasons rationally puts a constraint on his reasoning about the strategic situation. Rational reasoning entails that *i*'s set of consistent conditional probability systems can be determined by Definition 1 for some independent probability distribution $\pi_{-i}(\cdot)$ with a support that does not contain any vector of sets that *i* knows cannot be the vector of his opponents' actual best response sets.

The concept of rationalizable sets can now be defined and characterized.

DEFINITION 2. A non-empty set ρ_i is *rationalizable* for *i* if there exists $\omega \in \Omega$ with $\rho_i(\omega) = \rho_i$ such that it is common knowledge given ω that, $\forall i \in N$, *i* reasons rationally.

PROPOSITION 2. *A non-empty set* ρ_i *is rationalizable for i iff* $\exists \mathrm{P} \equiv \mathrm{P}_1 \times \ldots \times \mathrm{P}_n \subseteq \Sigma$ *with* $\rho_i \in \mathrm{P}_i$ *such that* $\mathrm{P} \subseteq \beta(\mathrm{P})$.

Proposition 2 states that ρ_i is a rationalizable set for *i* iff there exists an internally stable collection $\mathrm{P} \equiv \mathrm{P}_1 \times \ldots \times \mathrm{P}_n$ such that P_i contains ρ_i.

The following proposition establishes general existence and provides an algorithm.

PROPOSITION 3. *There exists,* $\forall i \in N$, *at least one rationalizable set for i. Write,* $\forall i \in N$, P_i^* *for the collection of rationalizable set for i. Then* $\mathrm{P}^* \equiv \mathrm{P}_1^* \times \ldots \times \mathrm{P}_n^*$ *is stable, and* $\beta^k(\Sigma)$ *converges to* P^* *in a finite number of iterations.*

The stability of P^* means that, $\forall i \in N$, ρ_i is a rationalizable set for *i* iff ρ_i is a best response set given some independent probability distribution in $\Delta(\mathrm{P}_{-i}^*)$. Moreover, P^* is the largest stable collection since, by Proposition 2, any stable collection is included in P^*. Finally, by Lemma 1 of Appendix A, the algorithm of Proposition 3 can be given the following decision-theoretic interpretation: Write, $\forall k \geq 1$, $\mathrm{P}^k \equiv \mathrm{P}_1^k \times \ldots \times \mathrm{P}_n^k := \beta^k(\Sigma)$. Then $\rho_i \in \mathrm{P}_i^k$ iff there exists $\omega \in \Omega$ with $\rho_i(\omega) = \rho_i$ such that, $\forall m = 0, \ldots, k-1$, [it is mutual knowledge that]^m, $\forall i \in N$, *i* reasons rationally.

4. CHARACTERIZATION OF BERNHEIM-PEARCE RATIONALIZABILITY

The present section shows that Bernheim-Pearce rationalizability can be completely characterized by imposing common knowledge of rational reasoning instead of common knowledge of rational choice. First, however, we have to define the ordinary best response operator.

DEFINITION 3. r_i is an *ordinary best response* to $p_{-i}(\cdot|\cdot)$ if, $\forall s_i \in S_i$, $u_i(r_i, p_{-i}(\cdot|S_{-i})) \geq u_i(s_i, p_{-i}(\cdot|S_{-i}))$.

By the finiteness of G, it follows that if $p_{-i}(\cdot|\cdot)$ is a conditional probability system, then there exists an ordinary best response to $p_{-i}(\cdot|\cdot)$. If $\pi_{-i}(\cdot)$ is an independent probability distribution in $\Delta(\Sigma_{-i})$, then $b_i^{ord}(\pi_{-i}(\cdot)) := \{r_i \in S_i | \exists\, p_{-i}(\cdot|\cdot)$ consistent with $\pi_{-i}(\cdot)$ such that r_i is an ordinary best response to $p_{-i}(\cdot|\cdot)\}$ denotes i's *ordinary* best response set. If P_{-i} is a nonempty rectangular subcollection of Σ_{-i}, let $\beta_i^{ord}(\mathrm{P}_{-i}) := \{ b_i^{ord}(\pi_{-i}(\cdot)) \,|\, \pi_{-i}(\cdot) \equiv \prod_{j \neq i} \pi_j(\cdot) \in \Delta(\mathrm{P}_{-i}) \}$. Say that a rationalizable set is *ordinary* if, $\forall i \in N$, rational reasoning for i given ω is defined by $\rho_i(\omega) \in \beta_i^{ord}(\mathrm{P}_{-i}^{\omega}(\omega))$.

PROPOSITION 4. *Let $R^* \equiv R_1^* \times \ldots \times R_n^*$ denote the set of rationalizable pure strategy profiles as defined by Bernheim (1984) and Pearce (1984). Then, $\forall i \in N$, $r_i \in R_i^*$ iff there exists an ordinary rationalizable set ρ_i for i such that $r_i \in \rho_i$.*

Hence, for each player i, the set of rationalizable pure strategies is equal to the union of i's ordinary rationalizable sets. There may exist multiple ordinary rationalizable sets for each player; this is the case in games with multiple strict Nash equilibria, since any strict Nash equilibrium constitutes a vector of ordinary rationalizable sets.

5. ALTERNATIVE BEST RESPONSE OPERATORS

The purpose of the present section is to combine common knowledge of rational reasoning with best response operators that generate cautious and/or sequentially rational behavior. First, we define three different, but related best response operators.

Let $p_i(\cdot|\cdot)$ be a conditional probability system and write $Y_{-i}^0 := S_{-i}$. Since G is finite, Y_{-i}^1, Y_{-i}^2, ... can be defined inductively by $Y_{-i}^k = Y_{-i}^{k-1} \setminus \text{supp}[p_{-i}(\cdot|Y_{-i}^{k-1})]$ for $k \in \{1,...,K\}$ such that $Y_{-i}^{K-1} \neq \varnothing$ and $Y_{-i}^K = \varnothing$. In the terminology of Blume et al. (1991), $(p_{-i}(\cdot|Y_{-i}^0),...,p_{-i}(\cdot|Y_{-i}^{K-1}))$ is a *lexicographic conditional probability system* with full support.

DEFINITION 4. r_i is an *admissible best response* to $p_{-i}(\cdot|\cdot)$ if, $\forall s_i \in S_i$, $(u_i(r_i, p_{-i}(\cdot|Y_{-i}^k)))_{k=0}^{K-1} \geq_L (u_i(s_i, p_{-i}(\cdot|Y_{-i}^k)))_{k=0}^{K-1}$.[4]

Following Mailath et al. (1993, Def. 2), the set $X \subseteq S$ is a *strategic independence (SI) for player i* in G, if (i) $X = X_i \times X_{-i}$, and (ii) $\forall s_i, t_i \in X_i$, $\exists r_i \in X_i$ such that $\forall s_{-i} \in X_{-i}$, $u_i(r_i, s_{-i}) = u_i(s_i, s_{-i})$ and $\forall s_{-i} \in S_{-i} \setminus X_{-i}$, $u_i(r_i, s_{-i}) = u_i(t_i, s_{-i})$. Write $H_i^* := \{X \subseteq S | X$ is a SI for player $i\}$ and $H_i^*(s_i) := \{X \in H_i^* | s_i \in X_i\}$.

DEFINITION 5. r_i is a *strategic independence respecting (SIR) best response* to $p_{-i}(\cdot|\cdot)$ if, $\forall X = X_i \times X_{-i} \in H_i^*(r_i) \setminus \{\varnothing\}$, $\forall s_i \in X_i$, $u_i(r_i, p_{-i}(\cdot|X_{-i})) \geq u_i(s_i, p_{-i}(\cdot|X_{-i}))$.

Let Γ be an extensive game without nature; with G being the corresponding PRNF. 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, $S(h) \equiv S_i(h) \times S_{-i}(h)$. Write $H_i^\Gamma := \{S(h) \subseteq S | h$ is an information set for i in $\Gamma\}$ and $H_i^\Gamma(s_i) := \{X \in H_i^\Gamma | s_i \in X_i\}$. A conditional probability system $p_i(\cdot|\cdot)$ determines for each information set h for i in Γ a conditional conjecture $p_{-i}(\cdot|S_{-i}(h))$.

---

[4] For two vectors a and b, $a \geq_L b$ iff whenever $b_k > a_k$, there exists $m < k$ such that $a_m > b_m$.

DEFINITION 6. Given an extensive game without nature Γ, r_i is a *sequential best response* to $p_{-i}(\cdot|\cdot)$ if, $\forall X_i \times X_{-i} \in H_i^{\Gamma}(r_i)$, $\forall s_i \in X_i$, $u_i(r_i, p_i(\cdot|X_{-i})) \geq u_i(s_i, p_{-i}(\cdot|X_{-i}))$.

The following propositions establish existence and provide characterizations for these best response operators.

PROPOSITION 5. *(i) If $p_i(\cdot|\cdot)$ is a conditional probability system, then there exists an admissible best response to $p_{-i}(\cdot|\cdot)$. (ii) If r_i is an admissible best response to $p_{-i}(\cdot|\cdot)$, then r_i is a SIR best response to $p_{-i}(\cdot|\cdot)$. (iii) If r_i is a SIR best response to $p_i(\cdot|\cdot)$, then, given any extensive game Γ with G as the corresponding PRNF, r_i is a sequential best response to $p_{-i}(\cdot|\cdot)$.*

PROPOSITION 6. *(i) There exists a conditional probability system $p_i(\cdot|\cdot)$ such that r_i is an admissible best response to $p_{-i}(\cdot|\cdot)$ iff r_i is not weakly dominated by a pure or a mixed strategy. (ii) There exists a conditional probability system $p_{-i}(\cdot|\cdot)$ such that r_i is a SIR best response to $p_i(\cdot|\cdot)$ only if r_i is not weakly dominated by a pure strategy. (iii) Given the extensive game Γ, there exists a conditional probability system $p_{-i}(\cdot|\cdot)$ such that r_i is a sequential best response to $p_{-i}(\cdot|\cdot)$ iff there exists in Γ a system of conjectures satisfying Bayes' law such that r_i is optimal at all of i's information sets that r_i does not preclude from being reached.*

Remarks: 1. The *if* part of Proposition 6(i) does not hold for G with $n > 2$ if the conditional probability system is required to be independent. *2.* There are examples of PRNF games where a SIR best response is weakly dominated by a *mixed* strategy. *3.* If G corresponds to an extensive game Γ, then $s_i \in S_i$ corresponds to a *plan of action* (Rubinstein, 1991) in Γ, precisely because G is a PRNF game. Hence, s_i does not specify actions in Γ at those of i's information sets that s_i precludes from being reached. This restriction to plans of action in extensive games is innocent when players do not make mistakes.

If $\pi_{-i}(\cdot)$ is an independent probability distribution in $\Delta(\Sigma_{-i})$, then $b_i^{adm}(\pi_{-i}(\cdot))$ / $b_i^{SIR}(\pi_{-i}(\cdot))$ / $b_i^{seq}(\pi_{-i}(\cdot))$ denote i's *admissible / SIR / sequential* best response set. If P_{-i} is a nonempty

rectangular subcollection of Σ_{-i}, let $\beta_i^c(\mathrm{P}_{-i}) := \{ b_i^c(\pi_{-i}(\cdot)) \,|\, \pi_{-i}(\cdot) \equiv \prod_{j \neq i} \pi_j(\cdot) \in \Delta(\mathrm{P}_{-i}) \}$ for c = *adm*, *SIR*, *seq*. Say that a rationalizable set is *admissible* if, $\forall i \in N$, rational reasoning for i given ω is defined by $\rho_i(\omega) \in \beta_i^{adm}(\mathrm{P}_{-i}^i(\omega))$. Say that a rationalizable set is *SIR* if, $\forall i \in N$, rational reasoning for i given ω is defined by $\rho_i(\omega) \in \beta_i^{SIR}(\mathrm{P}_{-i}^i(\omega))$. Say that a rationalizable set is *sequential* if, $\forall i \in N$, rational reasoning for i given ω is defined by $\rho_i(\omega) \in \beta_i^{seq}(\mathrm{P}_{-i}^i(\omega))$.

By the following proposition, the concept of admissible rationalizable sets refines Bernheim-Pearce rationalizability. Examples 2 and 5 of Section 6 illustrate that this refinement can be strict. The proposition holds also for SIR and sequential rationalizable sets.

PROPOSITION 7. *Let* $R^* \equiv R_1^* \times \ldots \times R_n^*$ *denote the set of rationalizable pure strategy profiles as defined by Bernheim (1984) and Pearce (1984). Then,* $\forall i \in N$, $r_i \in R_i^*$ *if there exists an admissible rationalizable set* ρ_i *for* i *such that* $r_i \in \rho_i$.

Börgers (1994) suggests combining admissibility with *approximate* common knowledge of rational choice. He shows that this leads to a procedure due to Dekel & Fudenberg (1990). In this procedure, which is also promoted by Gul (1995), first all weakly dominated strategies are eliminated, and then strictly dominated strategies are iteratively eliminated. The following proposition establishes that any strategy in an admissible rationalizable set survives this procedure.

PROPOSITION 8. *Let* $\widetilde{R} \equiv \widetilde{R}_1 \times \ldots \times \widetilde{R}_n$ *denote the set of pure strategy profiles surviving one round of elimination of weakly dominated strategies and then iterated elimination of strictly dominated strategies. Then,* $\forall i \in N$, $r_i \in \widetilde{R}_i$ *if there exists an admissible rationalizable set* ρ_i *for* i *such that* $r_i \in \rho_i$.

Examples 2 and 5 of Section 6 illustrate that not all strategies surviving this procedure are elements of some admissible rationalizable set. In these examples, combining admissibility with common knowledge of rational reasoning captures a notion of forward induction; combining admissibility with approximate common knowledge of rational choice does not achieve this.

6. EXAMPLES: FORWARD INDUCTION AND STRATEGIC MANIPULATION

In this section we present four games that are chosen to illustrate the notions of forward induction and strategic manipulation. For these games the concept of admissible rationalizable sets coincides with the concept of SIR rationalizable sets and—given the particular underlying extensive forms indicated in the text—with the concept of sequential rationalizable sets. The collection of admissible rationalizable sets is derived verbally in the discussion below and formally in Appendix B.

EXAMPLE 2: G_2 is the PRNF of an extensive form "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 the game to promote his extensive form rationalizability. Kohlberg & Mertens (1986) argue that the information contained in the PRNF G_2 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_2

In G_2 the analysis of the present paper supports the forward induction logic in the following manner: Since D is a strictly dominated strategy, and since 1 reasons rationally, D cannot be an element of 1's admissible best response set. Now, $\{R\}$ is 2's admissible best response set only if 2 assigns positive probability to $\{D\}$ or $\{U,D\}$ being 1's admissible best response set. Since 2 reasons rationally and knows that 1 reasons rationally, it follows that L has to be an element of 2's admissible best response set. This in turn implies that $\{U\}$ cannot be 1's admissible best response set. Knowing that $\{M\}$ or $\{U,M\}$ is 1's admissible best response set and reasoning rationally, player 2 will—conditional on 1 choosing from $\{M,D\}$—believe with

probability 1 that 1 is choosing M. Hence, common knowledge of rational reasoning implies that 2's admissible best response set is $\{L\}$, and, consequently, 1's admissible best response set is $\{M\}$. The argument above shows that $(\{M\},\{L\})$ is the unique vector of admissible rationalizable sets. The strategy profile implied by this vector entails that 1 can signal—by asking 2 to play—that he seeks a payoff as high as 2, leading to the implementation of 1's preferred B-o-S outcome.

It is noteworthy that common knowledge of rational reasoning combined with admissibility yields the forward induction outcome in G_2.[5]

EXAMPLE 3: G_3 is the PRNF of an extensive game due to Battigalli (1989) and Börgers (1991). Note that this game is a slight variation of G_2.

	L	***R***	
U	2,2	2,2	
M	0,1	3,0	
D	1,0	0,3	G_3

Let us investigate the consequences of imposing common knowledge of rational reasoning. Since D is a strictly dominated strategy, and since 1 reasons rationally, D cannot be an element of 1's admissible best response set. Now, $\{R\}$ is 2's admissible best response set only if 2 assigns positive probability to $\{D\}$ or $\{U,D\}$ being 1's admissible best response set. Since 2 reasons rationally and knows that 1 reasons rationally, it follows that L has to be an element of 2's admissible best response set. This in turn implies that $\{M\}$ cannot be 1's admissible best response set. However, common knowledge of rational reasoning cannot rule out that the remaining sets —$\{U\}$ and $\{U,M\}$ for 1 and $\{L\}$ and $\{L,R\}$ for 2—can be admissible best response sets. Hence, the collection of vectors of admissible rationalizable sets is $\{\{U\},\{U,M\}\}\times\{\{L\},\{L,R\}\}$.[6]

---

[5] Iterated elimination of weakly dominated strategies as well as Pearce's (1984) extensive form rationalizability and procedures proposed by Battigalli (1993a, 1993b) yield the forward induction outcome. In contrast to the present analysis, these procedures have not formally been given a common knowledge basis; however, see Stahl (1991).

[6] This means that in this example common knowledge of rational reasoning combined with admissibility yields a result that differs from that of the procedures listed in footnote 5.

It is instructive to verify that

$$\beta_1^{adm}(\pi_2(\cdot)) = \{U\} \text{ if } \pi_2(\{L\}) > 1/3 \text{ and } \pi_2(\{L,R\}) = 1 - \pi_2(\{L\})$$

$$\beta_2^{adm}(\pi_2(\cdot)) = \{U,M\} \text{ if } \pi_2(\{L\}) \leq 1/3 \text{ and } \pi_2(\{L,R\}) = 1 - \pi_2(\{L\})$$

$$\beta_2^{adm}(\pi_1(\cdot)) = \{L\} \text{ if } \pi_1(\{U\}) < 1 \text{ and } \pi_1(\{U,M\}) = 1 - \pi_1(\{U\})$$

$$\beta_2^{adm}(\pi_1(\cdot)) = \{L,R\} \text{ if } \pi_1(\{U\}) = 1.$$

To demonstrate that $\beta_1^{adm}(\pi_2(\cdot)) = \{U\}$ if $\pi_2(\{L\}) > 1/3$ and $\pi_2(\{L,R\}) = 1 - \pi_2(\{L\})$, note that a conditional probability system is consistent with $\pi_2(\cdot)$ iff 1's unconditional conjecture assigns at least $\pi_2(\{L\})$ probability to 2 playing L. It is now straightforward to verify that $\{U\}$ is the set of admissible best responses. To demonstrate that $\beta_2^{adm}(\pi_1(\cdot)) = \{L,R\}$ if $\pi_1(\{U\}) = 1$, note that a conditional probability system is consistent with $\pi_1(\{U\}) = 1$ if and only if 2 finds it infinitely more likely that 1 plays U rather than M or D. Hence, $p_1(U|S_1) = 1$ and $p_1(M|S_1) = p_1(D|S_1) = 0$. However, this puts no constraint on $p_1(\cdot|\{M,D\})$ and allows each of L and R to be an admissible best response to some conditional probability system consistent with $\pi_1(\{U\}) = 1$.

Observe that R is in 2's admissible best response set given $\pi_1(\{U\}) = 1$. This implies that 2 may—conditional on 1 choosing from $\{M,D\}$—assign positive probability to 1 choosing the strictly dominated strategy D. How is this compatible with Definition 1, which is based on the premise that 2 finds it infinitely more likely that 1 chooses a strategy in rather than outside his admissible best response set? To resolve this, note that if $\pi_1(\{U\}) = 1$, then 1 choosing from $\{M,D\}$ can be caused by each of the two following probability 0 events:

- Player 1's actual admissible best response set is not $\{U\}$.
- Player 1 chooses outside his admissible best response set.

Given $\pi_1(\{U\}) = 1$, Definition 1 does not exclude the possibility that player 2 will—conditional on 1 choosing from $\{M,D\}$—conclude that 1 chooses outside his admissible best response set, allowing 2 to assign positive probability to 1 choosing D.

In the extensive game of Battigalli (1989) and Börgers (1991) that G_3 represents, 1 and 2 move in sequence, with 2 being asked to play only if 1 does not choose the outside option U.

This extensive game can be called a game of *strategic manipulation*: Not to choose U is in 1's admissible rationalizable set only if he believes with sufficiently high probability that $\{L,R\}$ is 2's admissible best response set. Furthermore, $\{L,R\}$ is 2's admissible best response set only if she believes with probability 1 that $\{U\}$ is 1's admissible best response set. Hence, not to choose U can be explained only if 1 believes with sufficiently high probability that 2 believes with probability 1 that she cannot explain why he does not choose U. This argument also implies that not to choose U cannot be explained if 1 believes with sufficiently high probability that 2 believes with positive probability that she can explain why he does not choose U.

EXAMPLE 4: Dekel & Fudenberg (1990) consider an augmented version of G_2.

	L	***R***	***Z***	
U	2,2	2,2	½,½	
M	3,1	0,0	½,½	
D	0,0	1,3	½,½	G_4

G_4 is the PRNF of an extensive game where a "Battle-of-the-Sexes" game is preceded by opportunities for the players sequentially to exercise outside options: First player 2 may secure the payoffs (½,½). If she does not, player 1 may secure the payoffs (2,2). If none of this happens the B-o-S game is played. The derivation contained in Appendix B shows that the collection of vectors of admissible rationalizable sets is $\{\{M\},\{U,M\}\}\times\{\{Z\},\{L,Z\}\}$.

When Hammond (1993) discusses the underlying extensive game, he argues that there may be a tension between forward and backward induction in this game: For 2 to ask 1 to play may be interpreted as signaling that she seeks a payoff as high as ½, contrary to the payoff of 1 that 2 gets when the remaining subgame ($=G_2$) is considered an independent game. This may thereby induce 1 to play it safe by choosing U. The present analysis implies, however, that if 1 chooses U, he believes with probability 1 that 2 is making an inexplicable choice when she is not choosing Z. Hence, 'strategic manipulation' is a more appropriate term for such behavior by 2

than 'forward induction'. In particular, 2 hopes by asking 1 to play that 1 will choose U. However, if he does not, she will give in and play L. Also Dekel and Fudenberg (1990) consider (U,L) to be a reasonable outcome in this game.

EXAMPLE 5: G_5 is the PRNF of the "Burning Money" game discussed in the introduction.

	LL	***LR***	***RL***	***RR***	
NU	3,1	3,1	0,0	0,0	
ND	0,0	0,0	1,3	1,3	
BU	½,1	-½,0	½,1	-½,0	
BD	-½,0	-½,3	-½,0	-½,3	G_5

G_5 is the PRNF of a B-o-S game with the additional feature that player 1 can publicly destroy one and a half unit of utility 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 elimination of weakly dominated strategies) involves implementation of player 1's most preferred B-o-S outcome, with *no utility being burnt*. One might be skeptical about the iterated elimination of weakly dominated strategies in the "Burning Money" game because it seems to require 2 at one stage to judge burning by 1 as an explicable action, although burning eventually does not emerge as an explicable action.

As demonstrated in Appendix B, common knowledge of rational reasoning uniquely determines $\{NU\}$ as 1's admissible rationalizable set and $\{LL,LR\}$ as 2's admissible rationalizable set. Hence, the forward induction *outcome* is obtained, but 2 is free to interpret burning as she sees fit. This result follows from iterative elimination of *sets* of strategies, where at no stage of the iteration need 2 interpret burning as an explicable action since $\{NU\}$ is always included as a possible admissible best response set for 1.[7]

[7] Also Battigalli (1989), Asheim (1994), and Dufwenberg (1994) argue that (NU,LL) and (NU,LR) are the viable strategy profiles in the "Burning Money" game.

7. EXAMPLES: BACKWARD INDUCTION

During the last few years, a number of papers have discussed—in the context of extensive games—whether backward induction is implied by an assumption of common knowledge of rational choice.[8] The background for this interest is the following paradoxical aspect of backwards 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 mth stage of the game, the total pot is m dollars. If m is odd (even), player 1 (2) may take the m dollars and end the game, or leave it, in which case the pot increases with one dollar. 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. It is straightforward to show that in TOL(k), then with k odd (even), there are $(k+1)/2$ ($k/2$) admissible rationalizable sets for each player. Furthermore, for each player, the sets are nested. The smallest set contains only the backward induction strategy, while the largest set coincides with the set of strategies surviving one round of elimination of weakly dominated strategies and then iterated elimination of strictly dominated strategies. It is interesting to note that Aumann (1995) argues that common knowledge of rational choice leads to the former set of strategies, while Ben-Porath (1994) claims that common certainty[9] of rational choice leads to the latter set of strategies.

Binmore & Brandenburger (1990) observe that the backward induction paradoxes arise because players can "throw surprises" on one another by deviating from the backward induction path. Basu (1994) argues by way of an example (see below) that such backward induction paradoxes can occur also in a simultaneous move game. In the analysis of the present paper we have not distinguished between a simultaneous move game and an extensive game as long as the

---

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

[9] A player is certain of the event E if he assigns probability 1 to the event.

games have an identical PRNF. This implies that in the context of our analysis, it is inconsequential whether a backward induction outcome is based on *fait accompli* reasoning or based on *as if* reasoning. We therefore find it of interest to reconsider Basu's (1994) example.

EXAMPLE 6: In Basu's (1994) "Travelers' Dilemma" (TD) game, two players simultaneously announce bids, being integers between 2 and 100. If their bids coincide, each gets his bid. Else, the lowest bidder gets his bid +2, the other gets the lowest bidder's bid −2. In this game, (*2*,*2*) is the unique Bernheim-Pearce rationalizable strategy profile. Yet Basu argues that "there is something very rational about rejecting (*2*,*2*) and expecting your opponent to do the same". He draws a parallel between his game and extensive games in which backward induction paradoxes occur. However, he argues that the paradox in the TD game runs deeper because it is a simultaneous-move game in which players cannot "throw surprises on one other".

We show in Appendix B that in the TD game common knowledge of rational reasoning leads to a unique admissible rationalizable set for each player. This set contains only the backward induction strategy *2*. In TOL(k), player 2 need only optimize as if player 1 is not choosing the backward induction strategy. In the TD game, however, optimization uniquely determines a player's strategy if the opponent chooses the backward induction strategy. By modifying the TD game slightly so that a player need only optimize as if the opponent is not choosing the backward induction outcome, the results are drastically altered. Furthermore, we argue that this Modified TD (MTD) game more adequately serves to support Basu's (1994) intuition.

The MTD game has the same rules as the TD game *except that players are guaranteed a minimum payoff of 2 whatever they do.* In neither this game nor the TD game is it crucial that the highest bid is 100. Assume instead it is 4. Then we have

	2	*3*	*4*
2	2,2	4,0	4,0
3	0,4	3,3	5,1
4	0,4	1,5	4,4

TD: G_{6a}

	2	*3*	*4*
2	2,2	4,2	4,2
3	2,4	3,3	5,2
4	2,4	2,5	4,4

MTD: G_{6b}

These games have a lot in common. In particular, any subset not containing *2* is eventually eliminated by common knowledge of rational reasoning. But while the collection of vectors of admissible rationalizable sets is $\{\{2\}\}\times\{\{2\}\}$ for the TD game, for the MTD game it turns out that the collection of vectors of admissible rationalizable sets is $\{\{2\},\{2,3\}\}\times\{\{2\},\{2,3\}\}$. The reason why $\{2,3\}$ *is* included in the case of the MTD game, can be explained as follows. Suppose $\pi_{-i}(\cdot)$ satisfies $\pi_{-i}(\{2\}) = 1$. Then no constraints are imposed on i's conjecture conditional on $\{3,4\}$. Clearly, each of *2* and *3* can be an admissible best response to some conditional probability system consistent with $\pi_{-i}(\cdot)$. The reason why $\{2,3\}$ is *not* included in the case of the TD game, can be explained as follows. Given that $\pi_{-i}(\cdot) \in \Delta(\{\{2\},\{2,3\}\})$, for any $p_{-i}(\cdot|\cdot)$ consistent with $\pi_{-i}(\cdot)$, the *unique* ordinary best response to $p_{-i}(\cdot|\cdot)$ is *2*. It follows from Proposition 5(i) that this is also the unique admissible best response.

With 100 as the highest bid, the collection of vectors of admissible rationalizable sets is still $\{\{2\}\}\times\{\{2\}\}$ for the TD game. We find it hard to argue against this solution, even when the highest bid is large. In line with the interpretation of the algorithm of Proposition 3, $(\{2\},\{2\})$ as the unique vector of admissible rationalizable sets can be justified through unbounded iterated knowledge of rational reasoning. This iteration corresponds to a backward inductive (though *as if* rather than *fait accompli*) argument. By contrast, for the MTD game it turn out that $\{\{2\}, \{2,3\}, \ldots, \{2,3,\ldots,99\}\}\times\{\{2,\},\{2,3\}, \ldots, \{2,3,\ldots,99\}\}$ is the collection of vectors of admissible rationalizable sets if the highest bid is 100. We argue that *in this game* indecisiveness is plausible. If instead a unique vector of strategy sets were to be suggested, the only candidate would be $(\{2\},\{2\})$. However i's set of admissible best responses given $\pi_{-i}(\{2\}) = 1$ is $\{2,3,\ldots,99\}$. Hence, common knowledge of rational reasoning entails non-uniqueness in the MTD game.

Although we think that Basu's intuition is better supported by the MTD game than by the TD game, it is not unrealistic to predict that perfectly sensible humans will realize an outcome other than $(2,2)$ in actual play of the latter game. Still, we maintain that if there is common knowledge of rational reasoning, a player's admissible best response set will contain only *2*. However, in the MTD game his admissible best response set may contain more than *2*.

8. A FINAL REMARK

By Proposition 6(i), a strategy is an admissible best response to a conditional probability system iff it is dominated by neither a pure nor a mixed strategy. Hence, always choosing an admissible best response corresponds precisely to never choosing a weakly dominated strategy. The examples, however, show that admissible rationalizable sets do not generally coincide with the sets of strategies surviving iterated elimination of weakly dominated strategies. What can—in principle—account for these differences?

One difference occurs in games with three or more players since we impose that conditional probability systems be independent. This modeling choice—which leads a player to form plausible conjectures conditional on his opponents making inexplicable choices—is, however, inessential for the analysis.

The main distinguishing feature is that admissible rationalizable sets are arrived at by eliminating strategy *sets*, not by eliminating strategies. This has two consequences:

- Any strategy that is in the player's admissible best response set given some probability distribution over the opponents' remaining strategy sets is an available choice for the player.[10] When iteratively eliminating weakly dominated strategies, a strategy that is an admissible best response given the opponents' remaining strategies, may already have been eliminated.
- Through lexicographic optimization each player exploits opportunities for free insurance even against events that can occur only if opponents choose outside their remaining strategy sets. When iteratively eliminating weakly dominated strategies, no player exploits opportunities for free insurance against events that can occur only if opponents choose eliminated strategies.

We believe that both these consequences lend support to the approach to rationalizability proposed in the present paper.

---

[10] This follows since the algorithm of Proposition 3 determines a monotone sequence.

APPENDIX A: PROOFS

Proof of Prop. 1. *(i)* follows since, $\forall \pi_{-i}(\cdot) \equiv \prod_{j \neq i} \pi_j(\cdot) \in \Delta(\Sigma_{-i})$, $b_i(\pi_{-i}(\cdot)) \neq \varnothing$. *(ii)* $\Delta(P'_{-i}) \subseteq \Delta(P''_{-i})$. □

Proof of Prop. 2. *(If)* Let $E^P := \{\omega \in \Omega |\ \forall i \in N, \rho_i(\omega) \in P_i \text{ and } P'_{-i}(\omega) = P_{-i}\}$ and let $\Phi^P := \{\omega \in \Omega |\ \lim_{k \to \infty} \Pi^k(\{\omega\}) \subseteq E^P\}$. Then $\lim_{k \to \infty} \Pi^k(\Phi^P) \subseteq E^*$ since $P \subseteq \beta(P)$ implies that $E^P \subseteq E^*$. Finally, $\forall \rho_i \in P_i$, $\exists \omega \in \Phi^P$ such that $\rho_i(\omega) = \rho_i$. *(Only if)* Let $\omega \in \Omega$ satisfy $\rho_i(\omega) = \rho_i$ and $\lim_{k \to \infty} \Pi^k(\{\omega\}) \subseteq E^*$. Let, $\forall i \in N$, $P_i := \{\rho_i(\omega') |\ \exists k \geq 0$ with $\omega' \in \Pi^k(\{\omega\})\}$, and write $P \equiv P_1 \times \ldots \times P_n$. Then, $P \subseteq \beta(P)$ since, $\forall k \geq 0$, $\Pi^k(\{\omega\}) \subseteq E^*$. Finally, $\rho_i \in P_i$, since $\omega \in \Pi^0(\{\omega\})$. □

The epistemological analysis of Section 3 may be interpreted in terms of a *knowledge function*. Given the event $E \subseteq \Omega$, let KE denote the event that E is mutual knowledge. Hence, $KE := \{\omega \in \Omega |\ \Pi(\omega) \subseteq E\}$. Since, $\forall \omega \in \Omega$, $\omega \in \Pi(\omega)$, it follows that, $\forall E \subseteq \Omega$, $KE \subseteq E$.

Write, $\Phi^0 := \Omega$, and let, $\forall k \geq 1$, $\Phi^k := E^* \cap K\Phi^{k-1}$. I.e., $\forall k \geq 1$, $\Phi^k = \{\omega \in \Omega |\ \forall m = 0, \ldots, k-1$, [it is mutual knowledge that]^m, $\forall i \in N$, i reasons rationally$\}$. Let $\Phi^* := \{\omega \in \Omega |\ \lim_{k \to \infty} \Pi^k(\{\omega\}) \subseteq E^*\}$. I.e., $\Phi^* = \{\omega \in \Omega |$ it is common knowledge given ω that, $\forall i \in N$, i reasons rationally$\}$. Then it follows from the definitions that $\Phi^0 \supseteq \Phi^1 \supseteq \Phi^2 \supseteq \Phi^3 \supseteq \ldots \supseteq \Phi^*$.

LEMMA 1. *Let,* $\forall k \geq 0$, $P^k \equiv P_1^k \times \ldots \times P_n^k := \beta^k(\Sigma)$. *Then,* $\forall k \geq 0$, $\rho_i \in P_i^k$ *iff there exists* $\omega \in \Phi^k$ *with* $\rho_i(\omega) = \rho_i$.

Proof. (Only if) Repetitive use of Proposition 1(i)&(ii) implies that, $\forall k \geq 1$, $(\varnothing \neq) P^k \subseteq P^{k-1}$ $(\subseteq \Sigma)$. If $k=1$, then, $\forall \rho \in P^{k-1}$, $\exists \omega \in \Phi^{k-1}$ such that $\rho(\omega) = \rho$. Assume that this is true for some $k \geq 1$. Then $E^k := \{\omega \in K\Phi^{k-1} | \forall i \in N$, $\rho_i(\omega) \in P_i^k$ and $P'_{-i}(\omega) = P_{-i}^{k-1}\}$ is non-empty. Furthermore, since $P^k \subseteq \beta(P^{k-1})$ implies that $E^k \subseteq E^*$, it follows that $E^k \subseteq \Phi^k$. Finally, $P^k \subseteq P^{k-1}$ implies that, $\forall \rho \in P^k$, $\exists \omega \in E^k \subseteq \Phi^k$ such that $\rho(\omega) = \rho$. *(If)* Let, $\forall k \geq 0$ and $\forall i \in N$, $P_i^{\Phi^k} := \{\rho_i(\omega) |\ \omega \in \Phi^k\}$, and $P^{\Phi^k} \equiv P_1^{\Phi^k} \times \ldots \times P_n^{\Phi^k}$. Then $P^{\Phi^0} \subseteq P^0$. Assume that $P^{\Phi^{k-1}} \subseteq P^{k-1}$ for some $k \geq 1$. Then, since $\Phi^k \subseteq E^* \cap K\Phi^{k-1}$ implies that $P^{\Phi^k} \subseteq \beta(P^{\Phi^{k-1}})$, and $P^{\Phi^{k-1}} \subseteq P^{k-1}$ implies that $\beta(P^{\Phi^{k-1}}) \subseteq \beta(P^{k-1}) = P^k$, it follows that $P^{\Phi^k} \subseteq P^k$. □

LEMMA 2. *If* P^0, *with* $\varnothing \neq P^0 \subseteq \Sigma$, *is externally stable, then there exists* P^*, *with* $\varnothing \neq P^* \equiv P_1^* \times \ldots \times P_n^* \subseteq P^0$, *such that* P^* *is the largest (hence, unique maximal) internally stable collection included in* P^0.[11] *Furthermore,* P^* *is stable, and* P^k *defined by,* $\forall k \geq 1$, $P^k := \beta(P^{k-1})$, *converges to* P^* *in a finite number of iterations.*

Proof. Repetitive use of Proposition 1(i)&(ii) implies that, $\forall k \geq 1$, $(\varnothing \neq)\ P^k \subseteq P^{k-1}\ (\subseteq \Sigma)$. From this monotonicity and the finiteness of Σ, it follows that P^k converges in a finite number of iterations to $P^{k'}$ with $(\varnothing \neq)\ P^{k'} = \beta(P^{k'}) \subseteq P^0\ (\subseteq \Sigma)$. Let P^* denote the smallest rectangular collection that includes all internally stable collections included in P^0. There exists an internally stable collection included in P^0 since $P^{k'} \subseteq \beta(P^{k'}) \subseteq P^0$, and $P^{k'} \subseteq P^* \subseteq P^0$ since P^0 is rectangular. If P is an internally stable collection in P^0, then $P \subseteq \beta(P) \subseteq \beta(P^*)$ by Proposition 1(ii); i.e., $P^* \subseteq \beta(P^*)$ since $\beta(P^*)$ is rectangular. Hence, P^* is the largest internally stable collection included in P^0. As $(\varnothing \neq)\ P^{k'} \subseteq P^* \subseteq \beta(P^*) \subseteq \beta(P^0) = P^1\ (\subseteq \Sigma)$ (by Proposition 1(ii)), repetitive use of Proposition 1(ii) implies that, $\forall k \geq 1$, $(\varnothing \neq)\ P^{k'} \subseteq P^* \subseteq \beta(P^*) \subseteq \beta(P^{k-1}) = P^k\ (\subseteq \Sigma)$. Since P^k converges to $P^{k'}$, it follows that $P^{k'} = P^* = \beta(P^*)$. □

Proof of Prop. 3. Note that Σ is externally stable. By Proposition 2 and Lemma 2, P^* is the collection of vectors of rationalizable sets, where $P^* = \beta(P^*)$, and where $\beta^k(\Sigma)$ converges to P^* in a finite number of iterations. Repetitive use of Proposition 1(i)&(ii) implies that $\varnothing \neq P^* \subseteq \Sigma$. □

Proof of Prop. 5. Given $p_{-i}(\cdot|\cdot)$, write $Z_i^0 := S_i$ and define Z_i^1, Z_i^2, ... inductively by $Z_i^k := \arg\max_{s_i \in Z_i^{k-1}} u_i(s_i, p_{-i}(\cdot|Y_{-i}^{k-1}))$ for $k \in \{1,\ldots,K\}$. Then r_i is an admissible best response to $p_{-i}(\cdot|\cdot)$ iff $r_i \in Z_i^K$. *(i)* By the finiteness of G, $Z_i^K \neq \varnothing$. *(ii)* To show that $r_i \in Z_i^K$ implies that r_i is a SIR best response to $p_{-i}(\cdot|\cdot)$, suppose to the contrary that r_i is not a SIR best response to $p_{-i}(\cdot|\cdot)$. Then there exist $X = X_i \times X_{-i} \in H_i^*(r_i) \setminus \{\varnothing\}$ and $t_i \in X_i$ such that $u_i(r_i, p_{-i}(\cdot|X_{-i})) < u_i(t_i, p_{-i}(\cdot|X_{-i}))$. Since X is a SI, t_i can be chosen such that $\forall s_{-i} \in S_{-i} \setminus X_{-i}$, $u_i(t_i, s_{-i}) = u_i(r_i, s_{-i})$. Since

[11] A collection is *largest* if any other collection is included. A collection is *maximal* if no other collection strictly includes it.

$S_{-i} = Y_{-i}^0 \supset Y_{-i}^1 \supset \ \ldots \ \supset Y_{-i}^{K-1} \supset Y_{-i}^K = \varnothing$, there exists a largest integer $k \in \{1,\ldots,K\}$ satisfying $Y_{-i}^{k+1} \supseteq X_{-i}$; in particular it holds that $p_{-i}(X_{-i}|Y_{-i}^{k-1}) > 0$. By construction of t_i and Y_{-i}^{k+1}, either (a) both r_i and t_i are in Z_i^{k-1}, in which case it follows from Bayes' law $(\forall s_{-i} \in X_{-i},\ p_{-i}(s_{-i}|X_{-i}) \cdot p_{-i}(X_{-i}|Y_{-i}^{k-1}) = p_{-i}(s_{-i}|Y_{-i}^{k-1}))$ that $r_i \notin Z_i^k \supseteq Z_i^K$ since $\forall s_{-i} \in Y_{-i}^{k-1} \setminus X_{-i},\ u_i(r_i,s_{-i}) = u_i(t_i,s_{-i})$, or (b) both r_i and t_i are not in Z_i^{k-1}, in which case $r_i \notin Z_i^{k-1} \supseteq Z_i^K$. *(iii)* By Mailath et al. (1993, the *if* part of Theorem 1), $H_i^{\Gamma}(r_i) \subseteq H_i^*(r_i) \setminus \{\varnothing\}$. ∎

Proof of Prop. 6. (i) (If) By Pearce (1984, Lemma 4), if r_i is not weakly dominated by a pure or mixed strategy, there exists a conditional probability system $p_{-i}(\cdot|\cdot)$ with $p_{-i}(\cdot|S_{-i}) \in \Delta^0(S_{-i})$ such that, $\forall s_i \in S_i,\ u_i(r_i,p_{-i}(\cdot|S_{-i})) \geq u_i(s_i,p_{-i}(\cdot|S_{-i}))$. Since $K = 1$, r_i is an admissible best response to $p_{-i}(\cdot|\cdot)$. *(Only if)* Assume that r_i is weakly dominated by a (possibly degenerate) mixed strategy $p_i(\cdot) \in \Delta(S_i)$. By applying the notation of the proof of Proposition 5, it suffices to show that, for any conditional probability system $p_{-i}(\cdot|\cdot)$, $r_i \notin Z_i^K$. Note that $\{r_i\} \cup \text{supp}[p_i(\cdot)] \subseteq Z_i^0$. Furthermore, $\forall k \in \{1,\ldots,K\}$, $\{r_i\} \cup \text{supp}[p_i(\cdot)] \subseteq Z_i^{k-1}$ implies $(r_i \in Z_i^k$ only if $\text{supp}[p_i(\cdot)] \subseteq Z_i^k)$. To see this, observe that $u_i(r_i,p_{-i}(\cdot|S_{-i})) \leq \sum_{s_i \in S_i} p_i(s_i) u_i(s_i, p_{-i}(\cdot|Y_{-i}^{k-1}))$ since $p_i(\cdot)$ weakly dominates r_i. Also, if $r_i \in Z_i^k$, $\forall t_i \in Z_i^{k-1}$, $u_i(r_i,p_{-i}(\cdot|Y_{-i}^{k-1})) \geq u_i(t_i,p_{-i}(\cdot|Y_{-i}^{k-1}))$. Hence, $r_i \in Z_i^k$ implies, $\forall s_i \in \text{supp}[p_i(\cdot)] \subseteq Z_i^{k-1}$, $\forall t_i \in Z_i^{k+1}$, $u_i(r_i,p_{-i}(\cdot|Y_{-i}^{k-1})) = u_i(s_i,p_{-i}(\cdot|Y_{-i}^{k-1})) \geq u_i(t_i,p_{-i}(\cdot|Y_{-i}^{k-1}))$. However, $\{r_i\} \cup \text{supp}[p_i(\cdot)] \subseteq Z_i^K$ contradicts that $p_i(\cdot)$ weakly dominates r_i. *(ii)* Suppose, $\forall s_{-i} \in X_{-i},\ u_i(t_i,s_{-i}) > u_i(r_i,s_{-i})$ and, $\forall s_{-i} \in S_{-i} \setminus X_{-i},\ u_i(t_i,s_{-i}) = u_i(r_i,s_{-i})$. Then $\{r_i,t_i\} \times X_{-i}$ is a SI for i and there exists no conditional probability system $p_{-i}(\cdot|\cdot)$ such that r_i is a SIR best response to $p_{-i}(\cdot|\cdot)$. *(iii)* follows directly from the definition of a sequential best response. ∎

Let, $\forall i \in N,\ a_i(R_{-i}) := \{r_i \in S_i |\ \exists$ an independent conditional probability system $p_{-i}(\cdot|\cdot)$ with $p_{-i}(\cdot|S_{-i}) \in \Delta(R_{-i})$ such that r_i is an ordinary best response to $p_{-i}(\cdot|\cdot)\}$, where R_{-i} is a nonempty rectangular subset of S_{-i}. Write $a(R) \equiv a_1(R_{-1}) \times \ldots \times a_n(R_{-n})$, where $R \equiv R_1 \times \ldots \times R_n$. Note that $\varnothing \neq R' \equiv \prod_{i \in N} R_i' \subseteq R'' \equiv \prod_{i \in N} R_i'' \subseteq S$ implies $a(R') \subseteq a(R'')$. Following Pearce (1984), R satisfies the *best response property* if $R \subseteq a(R)$. By Bernheim (1984, Prop. 3.1) and Pearce (1984, Prop. 2), $R^* = a(R^*)$ is the largest subset of S satisfying the best response property.

Proof of Prop. 4. *(Only if)* $\mathrm{P} := \{(R_1^*,\ldots,R_n^*)\}$ satisfies $\mathrm{P} = \beta^{ord}(\mathrm{P})$ since $R^* = a(R^*)$. By Proposition 2, $\forall i \in N$, R_i^* is an ordinary rationalizable set for i. *(If)* By Proposition 3 and Definition 1, $\forall i \in N$, $R_i := \bigcup_{\sigma_i \in \mathrm{P}_i^*} \sigma_i = \bigcup_{\sigma_i \in \beta_i^{ord}(\mathrm{P}_{-i}^*)} \sigma_i \subseteq a_i(R_{-i})$. Since $R \subseteq a(R)$ implies $R \subseteq R^*$, it follows that, $\forall i \in N$, $\bigcup_{\sigma_i \in \mathrm{P}_i^*} \sigma_i \subseteq R_i^*$. □

Proof of Prop. 7. By Proposition 3 and Definition 1, $\forall i \in N$, $R_i := \bigcup_{\sigma_i \in \mathrm{P}_i^*} \sigma_i = \bigcup_{\sigma_i \in \beta_i^{adm}(\mathrm{P}_{-i}^*)} \sigma_i \subseteq a_i(R_{-i})$. Since $R \subseteq a(R)$ implies $R \subseteq R^*$, it follows that, $\forall i \in N$, $\bigcup_{\sigma_i \in \mathrm{P}_i^*} \sigma_i \subseteq R_i^*$. □

Proof of Prop. 8. Let $\widetilde{R}^1 \equiv \widetilde{R}_1^1 \times \ldots \times \widetilde{R}_n^1$ denote the set of pure strategy profiles surviving one round of elimination of weakly dominated strategies, and let, $\forall k > 1$, $\widetilde{R}^k \equiv \widetilde{R}_1^k \times \ldots \times \widetilde{R}_n^k$ denote the set of pure strategy profiles surviving, in addition, $k-1$ rounds of iterated elimination of strictly dominated strategies. Write $\mathrm{P}^0 = \Sigma$, and let, $\forall k \geq 1$, $\mathrm{P}^k \equiv \mathrm{P}_1^k \times \ldots \times \mathrm{P}_n^k := \beta^{adm}(\mathrm{P}^{k-1})$. By Proposition 6(i), $\forall i \in N$, $\bigcup_{\sigma_i \in \mathrm{P}_i^1} \sigma_i \subseteq \widetilde{R}_i^1$. Suppose, for some $k > 1$, $\forall i \in N$, $\bigcup_{\sigma_i \in \mathrm{P}_i^{k-1}} \sigma_i \subseteq \widetilde{R}_i^{k-1}$. Then, by Definition 1, $\forall i \in N$, $\bigcup_{\sigma_i \in \mathrm{P}_i^k} \sigma_i \subseteq a_i(\widetilde{R}_{-i}^{k-1}) \subseteq \widetilde{R}_i^k$. □

APPENDIX B: DERIVATIONS FOR THE EXAMPLES

The algorithm of Proposition 3 is used to determine the collection of vectors of admissible rationalizable sets in Examples 1-6.

Example 1.

$\mathrm{P}^0 = \Sigma = \quad \Sigma_1 \times \Sigma_2$
$\mathrm{P}^1 = \quad \{\{U\}\} \times \Sigma_2$
$\mathrm{P}^* = \mathrm{P}^2 = \quad \{\{U\}\} \times \{\{L\}\}$

Example 2.

$\mathrm{P}^0 = \Sigma = \quad \Sigma_1 \times \Sigma_2$
$\mathrm{P}^1 = \quad \{\{U\},\{M\},\{U,M\}\} \times \Sigma_2$
$\mathrm{P}^2 = \quad \{\{U\},\{M\},\{U,M\}\} \times \{\{L\},\{L,R\}\}$
$\mathrm{P}^3 = \quad \{\{M\},\{U,M\}\} \times \{\{L\},\{L,R\}\}$
$\mathrm{P}^4 = \quad \{\{M\},\{U,M\}\} \times \{\{L\}\}$
$\mathrm{P}^* = \mathrm{P}^5 = \quad \{\{M\}\} \times \{\{L\}\}$

Example 3.

$P^0 = \Sigma = \quad \Sigma_1 \times \Sigma_2$

$P^1 = \quad \{\{U\},\{M\},\{U,M\}\} \times \Sigma_2$

$P^2 = \quad \{\{U\},\{M\},\{U,M\}\} \times \{\{L\},\{L,R\}\}$

$P^* = P^3 = \quad \{\{U\},\{U,M\}\} \times \{\{L\},\{L,R\}\}$

Example 4.

$P^0 = \Sigma = \quad \Sigma_1 \times \Sigma_2$

$P^1 = \quad \{\{U\},\{M\},\{U,M\}\} \times \{\{R\},\{Z\},\{L,R\},\{R,Z\},\{L,Z\},\{L,R,Z\}\}$

$P^2 = \quad \{\{U\},\{M\},\{U,M\}\} \times \{\{Z\},\{L,R\},\{L,Z\}\}$

$P^3 = \quad \{\{M\},\{U,M\}\} \times \{\{Z\},\{L,R\},\{L,Z\}\}$

$P^* = P^4 = \quad \{\{M\},\{U,M\}\} \times \{\{Z\},\{L,Z\}\}$

Example 5.

$P^0 = \Sigma = \quad \Sigma_1 \times \Sigma_2$

$P^1 = \quad \{\{NU\},\{ND\},\{BU\},\{NU,ND\},\{ND,BU\},\{NU,BU\},\{NU,ND,BU\}\} \times \Sigma_2$

$P^2 = \quad \{\{NU\},\{ND\},\{BU\},\{NU,ND\},\{ND,BU\},\{NU,BU\},\{NU,ND,BU\}\} \times$
$\{\{LL\},\{RL\},\{LL,LR\},\{RL,RR\},\{LL,RL\},\{LL,LR,RL,RR\}\}$

$P^3 = \quad \{\{NU\},\{BU\},\{ND,BU\},\{NU,BU\},\{NU,ND,BU\}\} \times$
$\{\{LL\},\{RL\},\{LL,LR\},\{RL,RR\},\{LL,RL\},\{LL,LR,RL,RR\}\}$

$P^4 = \quad \{\{NU\},\{BU\},\{ND,BU\},\{NU,BU\},\{NU,ND,BU\}\} \times$
$\{\{LL\},\{RL\},\{LL,LR\},\{LL,RL\}\}$

$P^5 = \quad \{\{NU\},\{BU\},\{NU,BU\}\} \times \{\{LL\},\{RL\},\{LL,LR\},\{LL,RL\}\}$

$P^6 = \quad \{\{NU\},\{BU\},\{NU,BU\}\} \times \{\{LL\},\{LL,LR\},\{LL,RL\}\}$

$P^7 = \quad \{\{NU\},\{NU,BU\}\} \times \{\{LL\},\{LL,LR\},\{LL,RL\}\}$

$P^8 = \quad \{\{NU\},\{NU,BU\}\} \times \{\{LL\},\{LL,LR\}\}$

$P^9 = \quad \{\{NU\}\} \times \{\{LL\},\{LL,LR\}\}$

$P^* = P^{10} = \quad \{\{NU\}\} \times \{\{LL,LR\}\}$

Example 6a.

$P^0 = \Sigma = \quad \Sigma_1 \times \Sigma_2$

$P^1 = \quad \{\{2\},\{3\},\{2,3\}\} \times \{\{2\},\{3\},\{2,3\}\}$

$P^* = P^2 = \quad \{\{2\}\} \times \{\{2\}\}$

Example 6b.

$P^0 = \Sigma = \quad \Sigma_1 \times \Sigma_2$

$P^1 = \quad \{\{2\},\{3\},\{2,3\}\} \times \{\{2\},\{3\},\{2,3\}\}$

$P^* = P^2 = \quad \{\{2\},\{2,3\}\} \times \{\{2\},\{2,3\}\}$

REFERENCES

Asheim, G., 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.

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

Basu, K., 1994, "The Travelers' Dilemma", *American Economic Review* 84, Papers and Proceedings, 391–395.

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

Battigalli, P., 1993a, "On Rationalizability in Extensive Games", mimeo, Politecnico di Milano.

Battigalli, P., 1993b, "Strategic Rationality Orderings and the Best Rationalization Principle, Rapporto Interno #93.014, Politecnico di Milano.

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

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, 1007–28.

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

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

Binmore, K. and A. Brandenburger, 1990, "Common Knowledge and Game Theory", in *Essays on the Foundations of Game Theory*, 105–150, Oxford University Press.

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.

Cubitt, R. and R. Sugden, 1994, "Rationally Justifiable Play and the Theory of Non-Cooperative Games", *Economic Journal* 104, 798–803.

Dekel, E. and D. Fudenberg, 1990, "Rational Behavior with Payoff Uncertainty", *Journal of Economic Theory* 52, 243–67.

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

Greenberg, J., 1990, *The Theory of Social Situations*, Cambridge University Press.

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

Hammond, P., 1993, "Aspects of Rationalizable Behavior", in Binmore, Kirman, Tani (eds.), *Frontiers of Game Theory*, MIT Press.

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

Kohlberg, E. and P. Reny, 1992, "An Interpretation of Consistent Assessments", mimeo.

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.

Myerson, R., 1986, "Multistage Games with Communication", *Econometrica* 54, 323–358.

Pearce, D., 1984, "Rationalizable Strategic Behavior and the Problem of Perfection", *Econometrica* 52, 1029–50.

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., 1991, "Lexicographic Rationality, Common Knowledge, and Iterated Admissibility", Working Paper 91–10, Center for Economic Research, University of Texas.

Swinkels, J., 1994, "Independence for Conditional Probability Systems", mimeo.

van Damme, E., 1989, "Stable Equilibria and Forward Induction", *Journal of Economic Theory* 48, 476–96.