JOURNAL OF ECONOMIC THEORY **73**, 1–29 (1997)
ARTICLE NO. ET962229

Preference, Rationalizability and Equilibrium

Larry G. Epstein*

Department of Economics, University of Toronto, Toronto, Ontario M5S 3G7, Canada

Received September 22, 1995; revised April 17, 1996

In the context of finite normal form games, this paper addresses the formalization and implications of the hypothesis that players are rational and that this is common knowledge. The innovation is to admit notions of rationality other than subjective expected utility maximization. For example, rationality can be defined by the alternative restrictions that preferences are probabilistically sophisticated, conform to the multiple-priors model or are monotonic. The noted hypothesis is related to suitably defined notions of correlated rationalizability, survival of iterated deletion of strictly dominated strategies and a *posteriori* equilibrium. *Journal of Economic Literature* Classification Numbers: C72, D81.

1. INTRODUCTION

In the decision-theoretic approach to game theory, each player's problem of choosing a strategy is cast as a single agent decision problem under uncertainty. Then, assuming that players are Bayesian rational, alternative assumptions regarding their beliefs about the uncertainty that they face deliver axiomatizations of various solution concepts. An example of such an argument, that is the focus of this paper, is the theorem characterizing correlated rationalizability and survival of iterated deletion of strictly dominated strategies as the (equivalent) implications of rationality and common knowledge of rationality [28, Theorems 5.2–5.3]. These solution concepts are show in [10] to be equivalent in a suitable sense to *a posteriori* equilibrium, a streghtening of subjective correlated equilibrium [3]. Therefore, the assumptions of rationality and common knowledge of rationality also provide justification for this equilibrium notion.

As noted, in the received literature "rationality" is typically defined as Bayesian rationality; that is, each player forms a prior over the space of

* I gratefully acknowledge the financial support of the Social Sciences and Humanities Research Council of Canada and helpful comments from Peter Klibanoff, Bart Lipman, Michele Piccione, a referee, and especially Kin Chung Lo and Tan Wang. Eddie Dekel provided detailed comments on an earlier draft that lead to substantial improvements. This paper was written primarily during a very enjoyable visit at the Department of Economics and Faculty of Commerce, University of British Columbia.

0022-0531/97 $25.00

states of the world, identifies each strategy with an act over that state space, and maximizes the expected value of some vNM index. The objective of this paper is to describe a generalization of the noted results in which the definition of rationality is relaxed considerably. In particular, it will be necessary to provide appropriate definitions of "rationalizability", "dominance" and "*a posteriori* equilibrium" that are not tied to the subjective expected utility framework. (The term 'generalization' must be qualified. In common with [10], but unlike [28], this paper is restricted to *finite* normal form games. Therefore, references to [28, Theorems 5.2–5.3] should be interpreted as referring to the specializations of these results to finite games.)

There are three primary reasons for pursuing such a generalization. First, one objective of the decision-theoretic approach to game theory is to relate, at a formal level, our understanding of individual rationality on the one hand and strategic rationality on the other. But there remains a considerable gap or asymmetry in the formal modeling of rationality in the two settings. At the individual level, though subjective expected utility maximization is undoubtedly the dominant model in economics, many economists would probably view axioms such as transitivity or 'monotonicity' as more basic tenets of rationality than the Sure-Thing-Principle and other components of the Savage model. The implications of such more basic axioms for single agent decision-making are well understood from single agent abstract choice theory, but have they have not been isolated in strategic settings. One motivation for this paper is to narrow this gap. Second, the Ellsberg paradox and other evidence that people are averse to "ambiguity" or "vagueness" calls for a distinction between risk and uncertainty that is not possible within the Savage model. Under the presumption that uncertainty is important in strategic settings, concern with descriptive accuracy, therefore, calls for a notion of rationality that can accommodate such aversion to uncertainty. For example, rationality should not preclude conformity with axiomatic theories that have been developed in order to model uncertainty aversion, such as Choquet expected utility theory [26] or the multiple-priors model [15]. A final motivation is that within the more general framework provided here, the assumption that Bayesian rationality is common knowledge can be state formally, whereas it is well known that this assumption must be understood informally in the standard Bayesian framework.

A major difficulty in providing a formal analysis of the implications of rationality and common knowledge of rationality is the need to construct a state space that is a comprehensive representation of the uncertainty facing a given player, as is required of the space of states of the world. Such a construction is problematic because of the importance of 'beliefs about beliefs about beliefs...' and the resulting infinite regress. In the Bayesian framework, this difficulty has been resolved by [21] and [11], for example, whose constructions are employed in [28]. A corresponding foundation for the present paper is provided in [14], as summarized in Section 6 and Appendix A.

To conclude this introduction, it may be useful to provide perspective on this paper's contribution by acknowledging some of its limitations and by describing what is *not* attempted here. I have been emphasizing to this point that the implications of rationality and common knowledge of rationality depend on the formal definition of rationality. Naturally, they depend also on the definition of 'knowledge'. In this dimension, the paper follows the bulk of the literature by specifying (with a minor variation adopted for convenience) that an event is known if its complement is null in the sense of Savage. Consequently, strategies that are irrational for an opponent are given zero weight by any player contemplating her own strategy choice. It can be argued that total disregard for irrational play leads often to counterintuitive or empirically inaccurate predictions of play. Therefore, some authors have proposed theories of play consistent with a 'small' weight given to some forms of irrational play; see [7], [22] and [18], for example. These authors assume 'nonexpected utility' preferences—non-Archimedean expected utility in the former and forms of the multiple priors model in the latter two. But (when viewed from the present perspective) these generalized preferences are adopted primarily in order to better model alternative notions of knowledge rather than rationality. Another instance where the definition of knowledge is central is in the provision of decision-theoretic foundations for iterated deletion of weakly dominated strategies. This paper does *not* provide foundations for this deletion procedure. Preferences satisfying admissibility are allowed by the present framework and one could, in principle, restrict attention to such preferences. However, precisely as in the Bayesian case there is a contradiction between such 'full support' preferences and the knowledge, in the Savage sense, that some strategies are not played by the opponent. In other words, an alternative notion of knowledge is needed in order to justify the deletion of weakly dominated strategies. This would be the case also if our framework were expanded to admit lexicographic preferences; currently they are excluded, by the assumption that preferences have utility function representations, from the types space that is constructed in [14] and employed here. (See [4] and [27] for such use of lexicographic expected utility.) A final remark concerns the limited generality of the definition of rationality adopted here. Though the analysis weakens considerably the *a priori* restrictions on preferences beyond the subjective expected utility model, they are not eliminated entirely. In particular, the assumption that preferences have utility function representations presumes that they are transitive and complete. The latter assumption may be particularly troubling, because it might be argued that aversion to undertainty most naturally takes the form of incompleteness of preferences [6].

The paper proceeds as follows: Some preliminary notation and definitions are provided in Section 2. For the convenience of readers who may be interested in the new solution concepts and not necessarily in their 'foundations', the former are described first. Section 3 defines the generalized notion of (correlated) rationalizability for finite normal form games. The corresponding notion of 'dominance' is described in the following section. Next the equilibrium approach is studied. Section 6 employs the framework developed in [14] to show that the preceding solution concepts and equilibrium notion are characterized by the suitably specified assumption that players are rational and that this is common knowledge. Rationality is defined so as to be compatible not only with the Bayesian model. For example, it can accommodate any of the following alternative restrictions on preferences: ordinal expected utility [8], probabilistically sophisticated preferences [19], the multi-priors model [15] or monotonic preferences. These restrictions are defined precisely in Section 4.

2. PRELIMINARIES

This section introduces some notation and the key definition of a model of preference.

Consider a decision maker facing uncertainty represented by the state space S, a compact Hausdorff space. Objects of choice are acts over S, namely Borel measurable functions from S into the outcome space $[0, 1]$. The set of all such acts is denoted $\mathscr{F}(S)$. The universal class of preferences over $\mathscr{F}(S)$ is $\mathscr{P}(S)$, the class of *regular* preferences over $\mathscr{F}(S)$ as defined in Appendix A. For finite state spaces, preferences in $\mathscr{P}(S)$ are restricted roughly by the assumption that they admit representation by a utility function and by a weak monotonicity property. Even for general S, $\mathscr{P}(S)$ is a 'nonparametric' class, that is, it does not impose functional form assumptions such as expected utility. Each preference ordering in $\mathscr{P}(S)$ admits representation by a unique certainty equivalent utility function, so that we refer to elements of $\mathscr{P}(S)$ interchangeably as preference orderings or utility functions. Since a number of different state spaces arise below, it is convenient to view $\mathscr{P}(\cdot)$ as a correspondence on the domain $\mathscr{S}$ of nonempty compact Hausdorff state spaces that assigns $\mathscr{P}(S)$ to each $S \in \mathscr{S}$.

The 'knowledge' implicit in preferences will be important. Say that $u \in \mathscr{P}(S)$ *knows* the closed subset $E \subset S$ if the complement of E is null in the sense of Savage, that is, if any two acts that agree on $S \backslash E$ are ranked as indifferent by u. If E is not necessarily closed (or even measurable), say that *u knows E* if it contains a closed subset that is known by u. Use $\mathscr{P}(S \mid E)$

to denote the set of preferences that know E. Some readers may prefer the term 'believes E' rather than 'knows E'.

Two remarks are in order regarding this definition of knowledge. First, the use of closed subsets to define knowledge of an arbitrary set E is consistent with defining knowledge in a probabilistic setting by the condition that E contain the support of the relevant measure. (Recall that a support is closed by definition.) The weaker definition that E has probability 1 has a counterpart in our setting that involves simply replacing 'closed subset' by 'measurable subset' in the definition of 'u knows E'. We could adopt this weaker definition of knowledge here. Theorem 6.2 would remain valid for all models of preference that consist only of preferences that are continuous with respect to the norm topology on the space of acts (see [14, Section 5]). With this minor restriction, the remainder of the paper is unaffected. The two definitions of knowledge agree in the context of Sections 3–5, where only finite state spaces are relevant.

The second and more important remark is that alternative definitions of knowledge are possible and have appeared in the literature on the Choquet and multiple-priors models. In these frameworks, the definition of knowledge corresponds to the notion of 'support' that is used for the capacity or set of priors, and alternative notions have been proposed (see [13], [17] and [18], for example). When specialized to the Choquet and multiple-priors models, the above definition of 'know E' in terms of nullity of $S\backslash E$ implicitly adopts the least restrictive notion of support. An advantage of this definition is that it permits sharp results that can be interpreted as reflecting exclusively the more liberal meaning attached in this paper to rationality. On the other hand, the definition rules out concern by players with the possible irrationality of opponents, as explained in the introduction.

From the present perspective, the standard assumption in game theory that players are subjective expected utility maximizers corresponds to the restriction that players' preferences lie in a suitable subset of $\mathscr{P}(S)$. Here, we formalize alternative models of preference via alternative subsets of $\mathscr{P}(S)$, or more precisely via alternative subcorrespondences of $\mathscr{P}(\cdot)$.

We will be dealing with games involving two players i and j. Therefore, define a *model of preference* by a pair $\mathscr{P}^*(\cdot) \equiv (\mathscr{P}_i^*(\cdot), \mathscr{P}_j^*(\cdot))$, where $\mathscr{P}_k^*(\cdot)$ represents the admissible preferences for player k. Formally, $\mathscr{P}_k^*(\cdot)$ is a correspondence on $\mathscr{S}$ satisfying the conditions below. They make use of the following notation:

$$\mathscr{P}^*(S \mid S') \equiv \mathscr{P}(S \mid S') \cap \mathscr{P}^*(S), \qquad \forall S' \subset S \in \mathscr{S}, \text{ and} \tag{2.1}$$

for any $u \in \mathscr{P}(S \times S')$, $mrg_{\mathscr{F}(S)} u$ denotes the restriction of u to $\mathscr{F}(S)$, where the latter is identified in the natural way with a subset of $\mathscr{F}(S \times S')$.[1] The

[1] The product topology is used for all Cartesian product sets.

following conditions are understood to apply for all S and S' in $\mathscr{S}$ and $k=i,j$:

PREF1. $\varnothing \neq \mathscr{P}_k^*(S) \subset \mathscr{P}(S)$.

PREF2. Let $u \in \mathscr{P}_k^*(S)$, let σ: $S \to S'$ be continuous and satisfy *either* (a) σ is one-to-one, *or* (b) S and S' are finite. Then $u' \in \mathscr{P}_k^*(S')$ where $u'(f) \equiv u(f \circ \sigma)$.

PREF3. If $C \subset S$ is closed and $u \in \mathscr{P}_k^*(S \mid C)$, then $u' \in \mathscr{P}_k^*(C)$ where for any $f \in \mathscr{F}(C)$, $u'(f) = u(\hat{f})$, $\hat{f} = f$ on C and x on $S \backslash C$, $x \in [0,1]$.

PREF4. $\{mrg_{\mathscr{F}(S)} u\colon u \in \mathscr{P}_k^*(S \times S')\} \subset \mathscr{P}_k^*(S)$.

A preliminary comment on PREF2 is offered first. The either/or stipulation may seem unnatural. However, all assertions made in this paper regarding PREF2 are also true for the alternative assumption PREF2′, obtained by replacing '(a) or (b)' by the following condition: (c) The correspondence σ^{-1} admits a continuous selection, that is, there exists a continuous function ξ: $\sigma(S) \to S$ such that $\sigma(\xi(s')) \equiv s'$ on $\sigma(S)$. Each of (a) or (b) implies (c).[2] Therefore, PREF2′ is a *stronger* assumption than PREF2. The weaker PREF2 is adopted here. Note also that the interpretation of PREF2 that follows in this section refers to the case where σ is one-to-one. In the case of finite state spaces, 'regularity' of preferences or of measures is trivially satisfied and interpretation is clear.

The conditions PREF1–4 are readily verified for the expected utility model, for the other specific models of preference described in Section 4 and for the universal model $\mathscr{P}_k^*(\cdot) = \mathscr{P}(\cdot)$. Since verification is routine, no details will be provided. However, it might be useful to give a brief informal indication of their content for the expected utility model. In that model, if the vNM index is fixed, then each expected utility function can be identified with a unique regular (Borel) probability measure on the state space. In terms of this identification, the above conditions can be translated into the following largely familiar facts regarding regular probability measures: (PREF1) There exist regular probability measures on any compact Hausdorff space S; integration with respect to such a measure defines a functional on acts $\mathscr{F}(S)$ that is regular in the sense of (U.3, 4) of Appendix A. (PREF2) The regularity of a probability measure is preserved by a one-to-one and continuous transformation. (PREF3) Any regular probability measure m on S having support in $C \subset S$ can be viewed as a regular probability measure on C (PREF4). Given a regular probability measure on $S \times S'$, the S-marginal is regular on S. Roughly speaking, the formal notion of a model of preference is intended to capture counterparts of these properties

[2] In light of the compactness of S, σ one-to-one implies that it is open, that is, σ^{-1} is continuous.

appropriate for preferences on acts that are not necessarily integrals (expected values) with respect to some probability measure.

It is possible to further interpret and motivate PREF2–4. They restrict the way in which the 'admissible' class of preferences $\mathscr{P}_k^*(S)$ varies with S, roughly by ensuring that this class is sufficiently large. For PREF2, it is useful to consider four special cases. (i) If $S \subset S'$ and σ is the identity embedding, then PREF2 requires that if u is an admissible utility function for acts over the state space S, then the mapping $f \mapsto u(f_{|S})$ defines an admissibility utility function for acts over the larger state space S'; $f_{|S}$ denotes the restriction of f to S. (ii) In the special case of (i) where S is a singleton, we conclude that for each s' in S', the evaluation map $f \mapsto f(s')$ defines a utility function lying in $\mathscr{P}_k^*(S')$. (iii) Let $S' = \bar{S} \times S$ and $\sigma(s) = (\bar{s}, s)$ for some fixed $\bar{s}$. PREF2 requires that if u is an admissible utility function for acts over the state space S, then the mapping $f \mapsto u(f(\bar{s}, \cdot))$ must be an admissible utility function for the state space $\bar{S} \times S$. The latter utility function models knowledge that the first component of the state in S' is $\bar{s}$ and that the uncertainty associated with S is evaluated using u. (iv) If σ is onto, then it is a homeomorphism and PREF2 imposes the natural requirement that any admissible utility function on acts over S be transformed, by the homeomorphism, into an admissible utility function on acts over S'. PREF3 ensures that each admissible utility function in $\mathscr{P}_k^*(S)$ with 'support' contained in C can be identified with an admissible utility function for the state space C. PREF4 imposes that $\mathscr{P}_k^*(\cdot)$ behave in the natural way with respect to marginalization, in particular, that the marginal of any admissible preference for the product state space $S \times S'$ be an admissible preference for the component space S. In light of the implication (iii) of PREF2 just noted, if PREF2 is given, then PREF4 is *equivalent* to the set equality

$$\{mrg_{\mathscr{F}(S)} u \colon u \in \mathscr{P}_k^*(S \times S')\} = \mathscr{P}_k^*(S). \tag{2.2}$$

PREF2 is required for the proofs of Theorems 5.1 and 6.3. Together, PREF2 and PREF3 imply a one-to-one correspondence between $\mathscr{P}_k^*(C)$ and $\mathscr{P}_k^*(S \mid C)$, a property that is invoked in Section 4. Since $\mathscr{P}_k^*(C)$ is nonempty, so is $\mathscr{P}_k^*(S \mid C)$ as required in Sections 3 and 6. PREF4 renders Sections 3–5 compatible with Section 6. In fact, weaker conditions suffice in all cases. In particular, only finite state spaces are relevant in Sections 3–5, while the only infinite state spaces needed in Section 6 are subspaces of the space $A \times T$ described there.

3. RATIONALIZABILITY

Consider a two-player normal form game (A_i, A_j, r_i, r_j), where A_i and A_j are the finite strategy sets for players i and j, $r_i, r_j \colon A_i \times A_j \to X$ are outcome functions and the game is common knowledge. Here X denotes a compact interval in the real line. The extension to n players is immediate if "correlated rationalizability" is taken as the benchmark notion of rationalizability [10]. This paper will not consider the formulation of 'independence' needed to generalize rationalizability as defined in [5] and [23]. It is notationally simplifying to assume that $A_i = A_j = A$; the subscripts will be employed frequently where emphasis is desired.

The choice of strategy a_i by player i yields the uncertain outcome $r_i(a_i, \cdot)$ depending on j's choice of strategy. Thus the uncertainty faced by i is represented by the state space A_j and i's decision problem can be expressed in the following way: Each strategy a_i for i determines the act $r_i(a_i, \cdot)$ in $\mathscr{F}(A_j)$. Accordingly, i's strategy choice is determined by maximizing her preference ordering, an element of $\mathscr{P}(A_j)$. Similarly for j.

Note that explicit randomization is excluded; players choose elements of A, not probability distributions over A. This assumption is not innocuous, as illustrated in the multiple-priors example in Section 4.2, but is defensible on the grounds of the well-known conceptual difficulties surrounding explicit randomization and the often expressed view that people simply do not randomize when making decisions (see [24, Section 3] and [9, p. 91], for example). One of the noted conceptual difficulties arises from the fact that the commonly adopted expected utility framework excludes a strict incentive to randomize. Because much more general preferences are admitted here, reexamination is in order; that is, one might wonder whether at least some of those preferences imply a strict preference for randomization. The answer is that no such logical implication exists because: (i) the preferences postulated here are over Savage style acts; (ii) explicit randomization generates for each player a decision problem involving two-stage, horse-race/roulette-wheel or roulette-wheel/horse-race acts of the Anscombe–Aumann [1] variety, and (iii) there are ways of extending preferences to two-stage acts such that randomization is a matter of indifference, for example, if players perceive the game in such a way that they view their choice of mixed strategy as a choice between different roulette-wheel/horse-race acts (rather than the reverse order). Lo [17] makes this argument regarding 'perception of the game' and incentive to randomize in the context of a game with multiple-priors utility functions; his argument, in turn, is based on [12].

Fix a model of preference $\mathscr{P}^*$. For this section, the only important properties of $\mathscr{P}^*$ are that $\mathscr{P}^*(A)$ be defined for the specific finite state space A, that $\varnothing \neq \mathscr{P}^*_k(A) \subset \mathscr{P}(A)$, and that for all nonempty subsets $E \subset A$,

$$\mathscr{P}^*_k(A \mid E) \neq \varnothing. \tag{3.1}$$

The properties are implied by conditions PREF1–PREF3.

For any such model of preference $\mathscr{P}^*$, define a corresponding notion of rationalizability, called $\mathscr{P}^*$-rationalizability. Roughly, $\mathscr{P}^*$-rationalizability strategy profiles are such that player k's strategy is best response to some preference ordering that conforms to $\mathscr{P}_k^*(A)$ and for which the implicit beliefs about opponent's actions are 'justifiable'. More precisely, we adopt the following definition:

DEFINITION 3.1. The set of $\mathscr{P}^*$-*rationalizable* strategy profiles is the largest set $R_i \times R_j \subset A_i \times A_j$ with the property: For each $a_i \in R_i$ there exists a preference ordering u in $\mathscr{P}_i^*(A \mid R_j)$ such that $u(r_i(a_i, \cdot)) \geqslant u(r_i(a_i', \cdot))$ for all $a_i' \in A_i$; and similarly for j.

Under the conditions in the definition, say that a_i is a *best response* (*b.r.*) to u.

It is intuitive that the following iterative procedure delivers rationalizable strategy profiles. Let $R_i^0 = R_j^0 = A$ and for $n \geqslant 1$,

$$R_i^n = \{a_i \in A_i \colon a_i \text{ is a b.r. to some } u \in \mathscr{P}_i^*(A_j \mid R_j^{n-1})\}, \tag{3.2}$$

$$R_j^n = \{a_j \in A_j \colon a_j \text{ is a b.r. to some } u \in \mathscr{P}_j^*(A_i \mid R_i^{n-1})\}. \tag{3.3}$$

THEOREM 3.2. *The set of $\mathscr{P}^*$-rationalizable profiles $R_i \times R_j$ is nonempty and is given by*

$$R_i = \bigcap_{n=0}^{\infty} R_i^n \qquad \textit{and} \qquad R_j = \bigcap_{n=0}^{\infty} R_j^n. \tag{3.4}$$

Proof. Denote by R_i^∞ and R_j^∞ the two intersections in (3.4). Each R_i^n and R_j^n is nonempty because of (3.1). Note also that $R_i^n \searrow R_i^\infty$ and similarly for j. Therefore, if a_i is a b.r. to some preference in $\mathscr{P}_i^*(A_j \mid R_j^\infty) \subset \mathscr{P}_i^*(A_j \mid R_j^n)$, then $a_i \in R_i^n \; \forall n$, and so $a_i \in R_i^\infty$. Conversely, because the game is finite, $\exists l$ such that $R_j^l = R_j^{l+1} = \cdots = R_j^\infty$. (This implies that R_j^∞ is nonempty.) It follows that

$$R_i^\infty \subset \{a_i \in A_i : a_i \text{ is a b.r. to some } u \in \mathscr{P}_i^*(A_j \mid R_j^\infty)\};$$

and therefore set equality obtains. Similarly for j. Therefore, $R_i^\infty \times R_j^\infty \subset R_i \times R_j$. The reverse inclusion is obvious, because $R_i \times R_j \subset R_i^n \times R_j^n$ for every n. ∎

4. 'DOMINANCE' IN SPECIFIC MODELS

4.1. *Iterated Deletion*

In the Bayesian framework, rationalizability is equivalent to survival of iterated deletion of strictly dominated strategies. To establish a corresponding equivalence in this more general setting, one needs a characterization of strategies a_i that are not best responses in A_i to any $u_i \in \mathscr{P}_i^*(A_j)$. Call any such strategy dominated (given A_i and $\mathscr{P}_i^*(A_j)$). Theorem 3.2 suggests the iterated deletion procedure whereby all dominated strategies are eliminated from the initial game, the same for the resulting reduced game, and so on. It follows from the Theorem that this procedure delivers the $\mathscr{P}^*$-rationalizable strategy profiles. A complication is that beyond the first round, the relevant property is being a b.r. for some u_i in $\mathscr{P}_i^*(A_j \mid R_j^n)$ rather than in $\mathscr{P}_i^*(A_j)$. But as pointed out in Section 2, the former may be identified in a natural way with $\mathscr{P}_i^*(A_j \mid R_j^n)$. Therefore, the dominance characterization, with A_j replaced by R_j^n, applies at all stages of the deletion procedure.

The deletion procedure has the property that the surviving set of profiles is unaffected if only some, rather than all, dominated strategies are deleted at each round; that is, the order of deletion is of no consequence. As a result, some light may by shed on the conceptual issue of whether sensitivity to order of deletion, such as for weakly dominated strategies, seriously undermines a solution concept.[3] The pertinent question seems to be whether a solution concept that may be justified by being founded in common knowledge of rationality necessarily features insensitivity to order of deletion. Within the present framework, the answer is ‘yes’ providing a precise formal sense in which sensitivity to order of deletion reflects negatively on the solution concept.

That the set of $\mathscr{P}^*$-rationalizable strategy profiles is obtained regardless of the order of deletion follows from the following two facts: #1. If a_i is not a b.r. in A_i to any $u_i \in \mathscr{P}_i^*(A_j)$, then a_i is not a b.r. to any utility in $\mathscr{P}_i^*(A_j')$, where A_j' is any subset of A_j. #2. If neither a_i nor $a_i' \neq a_i$ is a b.r. in A_i to any $u_i \in \mathscr{P}_i^*(A_j)$, then a_i is not a b.r. in $A_i \backslash \{a_i'\}$ to any $u_i \in \mathscr{P}_i^*(A_j)$. These facts are readily proven. For #1, let $u_i \in \mathscr{P}_i^*(A_j')$. As noted in Section 2, the defining properties of a model of preference imply that $\mathscr{P}_i^*(A_j')$ may be identified with $\mathscr{P}_i^*(A_j \mid A_j') \subset \mathscr{P}_i^*(A_j)$. Therefore, u_i can be viewed as a utility function in $\mathscr{P}_i^*(A_j)$ and the hypothesis of #1 implies that a_i is not optimal for u_i. For #2, take any $u_i \in \mathscr{P}_i^*(A_j)$. By the finiteness of A_i, there exists $a^* \in A_i$ such that

$$u_i(r_i(a^*, \cdot)) > u_i(r_i(a_i, \cdot)), \qquad a^* \neq a_i'.$$

Therefore, a_i cannot be utility maximizing even if a_i' is deleted from A_i.

4.2. *Examples*

In this subsection, the notion of ‘dominance’ is characterized in explicit form for each of a number of noteworthy special models of preference. The

[3] Eddie Dekel suggested this ‘application.’

iterative procedure just outlined can then be used to deliver the appropriate sets of rationalizable strategy profiles. These special cases clarify the way in which $\mathscr{P}^*$-rationalizability generalizes the familiar expected-utility-based notion or rationalizability.

Incomplete specifications of the various models of preference are provided in that only $\mathscr{P}^*(A)$ is defined in each case. Similar specifications apply to any other finite state space. For infinite state spaces, some further technical details are needed as described in Section 6.

The game defined in Fig. 1 will be used to illustrate the differing implications of $\mathscr{P}^*$-rationalizability for the various specifications of $\mathscr{P}^*$. The row player is i and the column player is j. The indicated payoff to the strategy pair (a_i, a_j) is $r_i(a_i, a_j)$ for i and similarly for j.

Analysis of the game is straightforward. If j is justified in assuming that i will not play M, then she will play R. If i believes this, she will play T. As a result, only (T, R) will be rationalizable. On the other hand, if M is a best response for i to a utility function in the class under consideration, then all strategy combinations will be rationalizable. Thus the set of rationalizable profiles is either (i) $\{(T, R)\}$ or (ii) all profiles. Assume for the moment that payoffs are in utility units, as explained in the next subsection. Then the standard expected utility model implies that (T, R) alone is rationalizable for all admissible parameter values (because $\delta < 1$). Intuitively it may seem problematic to exclude (M, L) as a rational play, particularly with the size of γ unrestricted. On the other hand, all generalizations of expected utility to be described include (M, L) in the rationalizable set at least for some (differing) parameter values, as indicated in the accompanying Table. (The assertions can be proven using the characterizations of dominance provided for each model of preference. Alternatively, for this simple game, a diagrammatic technique described below can be used. The reader may wish to refer to the Table while reading the examples.)

Expected utility. This is the standard model. For each player $k = i, j$, fix a continuous and strictly increasing vNM index v_i: $X \to \mathscr{R}$. For each p in

	L	R
T	$1+\gamma-\delta$, .9	2, 1
M	$1+\gamma$, 100	1, 1
B	$2+\gamma$, .9	$1-\delta$, 1

FIG. 1. Illustrative game, $0 \leqslant \delta < 1$, $0 \leqslant \gamma$.

$M(A)$, the set of probability measures on A, define the certainty equivalent function $u_k(\cdot; p) \in \mathscr{P}(A)$ by

$$u_k(f; p) = v_k^{-1}\left(\int_A v_k(f)\, dp\right).$$

Then define $\mathscr{P}_k^{EU}(A) = \{u_k(\cdot; p) : p \in M(A)\}$ and let $\mathscr{P}^{EU}(A)$ be the corresponding profile of preference classes. (The dependence on the vNM index v_k is suppressed in the notation as it is in similar situations below.) For this specialization of $\mathscr{P}^*$, $\mathscr{P}^*$-rationalizability reduces to the usual notion due to [5] and [23].

In addition, there is the well-known equivalence between rationalizability and survival of iterated deletion of strictly dominated strategies, where dominance is defined so as to include dominance by mixed strategies. This equivalence is due to the equivalence between (1) a_i not being strictly dominated, and (2) a_i being a best response for some beliefs over the relevant state space A.

Ordinal expected utility. The preceding models the situation where beliefs of players are not common knowledge but their vNM indices are common knowledge. To reduce this asymmetry, Borgers [8] assumes that only preferences over pure strategy outcomes of the game are common knowledge. Here, pure strategy outcomes are real numbers that are ranked in the usual way. Therefore, Borgers' assumption is captured by the following specification for $\mathscr{P}^*$: For any (strictly increasing) vNM index v as above, define

$$u(f; p; v) = v^{-1}\left(\int_A v(f)\, dp\right),$$

and let $\mathscr{P}_k^{OEU}(A) = \{u(\cdot; p; v) : p \in M(A),\ v \text{ a vNM index}\}$ for each player k.

Borgers shows that $a^* \in A_i$ is not a best response to any $u_i \in \mathscr{P}_i^{OEU}(A_j)$ *if and only if*: $\forall B_j \subset A_j\ \exists a_i$ such that

$$r_i(a_i, \cdot) \geqslant r_i(a^*, \cdot) \text{ on } B_j \qquad \text{and} \qquad \exists a_j \in B_j \text{ s.t. } r_i(a_i, a_j) \neq r_i(a^*, a_j). \quad (4.1)$$

Refer to strategies a^* satisfying this condition as *OEU-dominated* though dominance so defined does not correspond to a binary relation on i's strategy set. Iterated deletion of such dominated strategies leaves only $\mathscr{P}^{OEU}$-rationalizable strategy profiles.

Probabilistic sophistication. Drop the assumption of an expected utility functional form but continue to assume that players' preferences are based on probabilistic beliefs in the sense of [19]. Machina and Schmeidler refer to such preferences as 'probabilistically sophisticated.' Consideration of this class of preferences is motivated in part by the desire to accommodate behavioral evidence, such as the Allais paradox, contradicting the von Neumann–Morgenstern model of choice amongst risky prospects, that is, objective lotteries.

To describe probabilistically sophisticated preferences, adopt the following notation: For any act $f \in \mathscr{F}(A)$ and probability measure $p \in M(A)$, denote by $F_{p,f}$ the induced cumulative distribution function (cdf) on X. The set of all cdf's is $D(X)$. A function $V: D(X) \to X$ is called a *risk certainty equivalent* (r.c.e.) if it is strictly increasing in the sense of first degree stochastic dominance and if $V(F_{p,f}) = x$ for $f(\cdot) \equiv x$. Then the class of probabilistically sophisticated preferences is, for each player k,

$$\mathscr{P}_k^{ps}(A) = \{u \in \mathscr{P}(A) \colon u(f) \equiv V(F_{p,f}), \text{ for some } p \in M(A) \text{ and r.c.e. } V\}. \tag{4.2}$$

The relaxation from the ordinal expected utility class to probabilistically sophisticated preference does not change the set of rationalizable strategy profiles and in that sense has no empirical significance. To see this, note that if a_i is a b.r. to some $u \in \mathscr{P}_i^{ps}(A_j)$, then it violates OEU-dominance condition (4.1) for the set B_j defined as the support of the measure p underlying u in the sense of (4.2). It follows that a_i is also a b.r. to some $u \in \mathscr{P}_i^{OEU}(A_j)$. In other words, the property of being a best response to some admissible preference is *equivalent* whether one uses the ordinal expected utility class or the larger one consisting of all probabilistically sophisticated preferences.

Multiple-priors utility. Next consider a generalization of the standard expected utility model due to [15] that is motivated by the desire to model uncertainty aversion such as is exhibited in the Ellsberg paradox. In particular, as such aversion contradicts probabilistic sophistication, this class of preferences violates probabilistic sophistication. Note that this model is an alternative to, rather than a generalization of, the two preceding models.

Fix vNM indices v_k and for each closed and convex set of probability measures Δ on A, define $u_k(\cdot; \Delta) \in \mathscr{P}(A)$ by

$$u_k(f; \Delta) = v_k^{-1}\left(\min_{p \in \Delta} \int_A v_k(f)\, dp\right). \tag{4.3}$$

The multiple-priors class of preferences for player k is $\mathscr{P}_k^{MP}(A) = \{u_k(\cdot; \Delta) \in \mathscr{P}(A) \colon \Delta \text{ varies as above}\}$.

The problem to be addressed is the characterization of strategies $a^* \in A_i$ that are not best responses to any $u \in \mathscr{P}_i^{MP}(A_j)$. Refer to a^* as *MP-dominated* if it is never a best response. In Appendix B we derive the following characterization of such dominance: $\exists\{\alpha_a, \beta_a\}_{a \in A_i^*} \subset [0, 1]$, $\sum \alpha_a = \sum \beta_a = 1$, $A_i^* \equiv A_i \backslash \{a^*\}$, such that

$$\sum_{a \in A_i^*} \left(\min_{\alpha_j} \left[\alpha_a g_a(a_j) - \beta_a g_{a^*}(a_j)\right]\right) > 0, \tag{4.4}$$

where $g_a(a_j) \equiv v_i(r_i(a, a_j))$ describes the 'act' with utility payoffs facing player i when she chooses strategy a.

Comparison with more familiar dominance notions should help to clarify (4.4). By *pure strategy dominance* of a^*, we mean that

$$\exists a \in A_i \text{ s.t. } r_i(a, a_j) > r_i(a^*, a_j) \qquad \forall a_j \in A_j, \tag{4.5}$$

or equivalently (by monotonicity of v_i), $\exists a \in A_i$ such that

$$g_a(a_j) \equiv v_i(r_i(a, a_j)) > v_i(r_i(a^*, a_j)) \equiv g_{a^*}(a_j) \qquad \forall a_j \in A_j. \tag{4.6}$$

It is easy to see that pure strategy dominance of a^* implies that it is MP-dominated (take $\alpha_a = \beta_a = 1$ for the strategy a satisfying (4.6)). But the converse is not true in general; consider the illustrative game with parameters $\gamma > \delta + \delta^2$.

Comparison with the usual notion of (mixed strategy) strict dominance is also straightforward.[4] The minimum operator in (4.4) can be taken outside the summation without affecting the inequality. This yields the strict dominance of $\sum \alpha_a g_a$ over g_{a^*}. The converse is false; let $\gamma \leqslant \delta + \delta^2$ in the illustrative game. (Klibanoff [16] shows, in contrast, that equivalence obtains between survival of iterated deletion of dominated (in the usual sense) strategies and multiple-priors rationalizability if the latter is defined to admit the use of mixed strategies. We discussed earlier a rationale for excluding mixed strategies. There is no censensus in the related literature regarding the use of mixed strategies; mixed strategies matter in [17], but they play no role in [13], [22] and [18].)

Finally, iterated deletion based on (4.4) and Theorem 3.2 delivers the $\mathscr{P}^{MP}$-rationalizable strategy profiles. Beyond the first round, the relevant property is being a b.r. for u_i in $\mathscr{P}_i^{MP}(A_j \mid R_j^n)$. But the latter may be identified in the natural way with $\mathscr{P}_i^{MP}(R_j^n)$; therefore, the characterization (4.4), with A_j replaced by R_j^n, applies at all stages of the deletion procedure.

[4] Other comparisons can be made. First, OEU-dominance implies strict dominance by mixtures and therefore also MP-dominance. Second, $a \in A_i$ 'dominates' a^* according to the maximim criterion in the special case of (4.4) having $\alpha_a = \beta_a = 1$. Therefore, MP-dominance is more demanding and accordingly leads to the deletion of fewer strategies than does the maximin criterion.

ε-Contamination. The multiple-priors model of preference leaves the set of priors of each player unrestricted except by technical conditions, yielding a model that may be too general for some settings. Here consider a family of models that is parametrized by a single parameter ε such that $\varepsilon=0$ yields the expected utility model, $\varepsilon=1$ yields the multiple-priors model, and intermediate models are implied for $0<\varepsilon<1$. In cases where the modeler is willing to take a stand on the size of ε, a set of rationalizable strategy profiles smaller than in the preceding example may be obtained.

For each ε in $[0, 1]$, say that the set of priors Δ on A is an ε-contaminated set if there exists a probability measure $p^* \in M(A)$ and a closed and convex set C of probability measures on A such that[5]

$$\Delta = \{(1-\varepsilon)\, p^* + \varepsilon p : p \in C\}.$$

Denote by Υ^{ε} the set of all ε-contaminated sets of priors. One might think of p^* as 'benchmark' Bayesian prior, and that there exists some uncertainty or ambiguity that is captured by admitting the indicated contaminations of p^* with measures in C. If all conceivable contaminations are admitted, $C=M(A)$, one obtains the ε-contamination model that has received attention in the robust statistics literature. In the general case, the set Δ increases in the sense of set inclusion as $\varepsilon \nearrow$ with p^* and C fixed, modeling increased uncertainty aversion. For another perspective on ε, note that

$$\sup_{m, m' \in \Delta} \sup_{E \subset A} |m(E) - m'(E)| \leqslant \varepsilon. \tag{4.7}$$

Therefore, a small value for ε limits the differences between measures in Δ and indicates, in a natural way, a small difference from the single prior case.[6]

[5] As there is a superficial similarlity between this model and Mukerji's [22], this may be an appropriate place to clarify further the relationship between the two papers. Roughly, Mukerji assumes the following: Suppose that, in the context of an inductive procedure, a set $A_j^0 \subset A_j$ of 'rational' strategies for j has been determined. Then it is assumed that i, in evaluating her strategy opetions, has multiple-priors beliefs given by $\Delta = (1-\varepsilon)\, p^* + \varepsilon M(A_j)$, where p^* is a probability measure on A_j^0 and ε is a parameter in $[0, 1]$ giving the weight attached by i to j being irrational and 'therefore' choosing some unrestricted strategy in A_j. Mukerji deliberately avoids imposing knowledge of rationality in the sense of this paper, which is imposed here by replacing the contaminating set $M(A_j)$ by $M(A_j^0)$ (or any subset $C \subset M(A_j^0)$). In his model, uncertainty is due entirely to uncertainty about whether or not the opposing player is rational, whereas here rationality is assumed known but there is uncertainty about which rational strategy the opponent will play. As a result, neither model is nested in the other.

[6] One might use (4.7) directly rather than the ε-contamination specification to define sets of priors. However, the linearity in the ε-contamination structure is crucial for the derivation of the 'dominance' characterization that follows. On the other hand, specifications other than the ε-contamination structure can also be handled. Similar arguments can be used for any other specialization of the multiple-priors class that takes the form of restricting the set of priors Δ by a finite set of linear inequality restrictions on the component measures.

Define the model $\mathscr{P}^{\varepsilon}$ exactly like $\mathscr{P}^{MP}$ except that the sets of priors Δ are restricted to lie in Υ^{ε}. The arguments for the multiple-priors case are readily extended (see Appendix B) to prove that a^* is not a b.r. to any $u \in \mathscr{P}_i^{\varepsilon}(A_j)$ if and only if a^* is *ε-dominated*, where the latter means the following (using the notation from (4.4)): $\exists \{\alpha_a, \beta_a\}_{a \in A_i^*}$ in $[0, 1]$, $\sum \alpha_a = \sum \beta_a = 1$, such that

$$(1-\varepsilon) \min_{a_j} \left(\sum_{a \in A_i^*} \alpha_a g_a(a_j) - g_{a^*}(a_j) \right) + \varepsilon \sum_{a \in A_i^*} \min_{a_j} \left[\alpha_a g_a(a_j) - \beta_a g_{a^*}(a_j) \right] > 0. \tag{4.8}$$

This dominance notion is a 'convex combination' of mixed strategy strict dominance (corresponding to the first term) and *MP*-dominance (corresponding to the second term). Condition (4.7) may provide a basis for selecting a value for ε and accordingly the stringency of the dominance condition to be applied in any particular setting.

Monotonicity. Refer to $u \in \mathscr{P}(A)$ as *monotonic* if for all acts

$$f'(\cdot) > f(\cdot) \text{ everywhere on } A \text{ implies } u(f') > u(f).$$

For each player k, define $\mathscr{P}_k^{mon}(A)$ as the set of all monotonic utilities in $\mathscr{P}(A)$. This model of preference strictly generalizes the preceding models because the latter impose not only monotonicity but also various functional form restrictions on the representing utility functions. For example, the expected utility functional form is assumed in the ordinal expected utility class $\mathscr{P}_k^{OEU}(A)$ and probabilistically sophisticated preferences evaluate acts via induced probability distributions.

A natural conjecture is that a_i is a b.r. for some $u_i \in \mathscr{P}_i^{mon}(A_j)$ if and only if a_i is *not* strictly dominated in the pure strategy sense (4.5). Clearly, if a_i is so dominated, then it cannot be a best response. On the other hand, if it is undominated, it is a best response to $u \in \mathscr{P}_i^{mon}(A_j)$ defined by: For each f, an act over A_j,

$$u(f) \equiv \min_{a_j \in A_j} \{ f(a_j) - r_i(a_i, a_j) \} + \max_{a_j \in A_j} r_i(a_i, a_j).$$

In particular, $u(r_i(a, \cdot)) \leqslant u(r_i(a_i, \cdot)) = \max_{a_j \in A_j} r_i(a_i, a_j)$ for all $a \in A_i$. Conclude that $\mathscr{P}^{mon}$-rationalizability is equivalent to survival of the iterated deletion of strictly dominated strategies, where dominance is by pure strategies.

TABLE I

$\mathscr{P}^*$-Rationalizable Strategies ($\tau_\varepsilon \equiv \min\{\varepsilon(1+\delta)+\delta-1, \varepsilon\delta(1+\delta)\}$)

	$\mathscr{P}^{EU}$	$\mathscr{P}^{OEU}$	$\mathscr{P}^{ps}$	$\mathscr{P}^{MP}$	$\mathscr{P}^{\varepsilon}$	$\mathscr{P}^{mon}$
(T, R) only	$\delta<1$	$\delta=0$	$\delta=0$	$\gamma>\delta+\delta^2$	$\varepsilon\gamma>\tau_\varepsilon$	—
All (a_i, a_j)	—	$\delta>0$	$\delta>0$	$\gamma\leqslant\delta+\delta^2$	$\varepsilon\gamma\leqslant\tau_\varepsilon$	All (δ, γ)

4.3. *Summary*

It may be useful to reflect briefly on the differing implications of the alternative solution concepts as presented in Table I for the illustrative game. First, a diagrammatic derivation of these implications is described.[7] It is convenient in this derivation to assume that payoffs are denominated in the units of vNM utility indices. This may be accomplished for the models $\mathscr{P}^{EU}$, $\mathscr{P}^{MP}$ and $\mathscr{P}^{\varepsilon}$ by assuming that $v_k(x) \equiv x$ in each case. For the remaining models, the units in which payoffs are denominated is of no consequence.

Focus on whether M is a best response for i for some preference in the model defined by $\mathscr{P}^*$. Because there are only two feasible strategies for j, the choice for i between T, M and B can be portrayed as the choice between the three 'acts' portrayed in the plane in Fig. 2. The question then is whether one can draw an indifference curve through M that is consistent with the model $\mathscr{P}^*$ and that passes (weakly) above both T and B. The answer depends on the particular model $\mathscr{P}^*$, because different models restrict indifference curves in different ways.

The ordinal expected utility model can be fit into this framework by viewing it as attaching ordinal (rather than cardinal or absolute) significance to utility payoffs. In other words, any ordinal transformation of these payoffs is permitted, leaving indifference curves downward sloping but not necessarily linear. More precisely, the dominance condition (4.1) can be translated into the class of admissible indifference curves that are either (i) strictly downward sloping everywhere, or (ii) perfectly horizontal, or (iii) perfectly vertical. (The latter two cases correspond to beliefs by i that attach 0 probability to one of j's strategies.) It is therefore apparent that M can be a best response for any $\delta>0$. Intuitively, M is preferable to T if j plays L and it is preferable to B if she plays R. In both cases the gain in utility units is δ and δ can be 'large' after rescaling utilities. In other words, after a suitable rescaling of utilities, we are back in the standard model with $\delta=1$, where M is a best response. A similar discussion applies to the model of probabilistically sophisticated preferences.

[7] I am grateful to Kin Chung Lo for suggesting this technique.

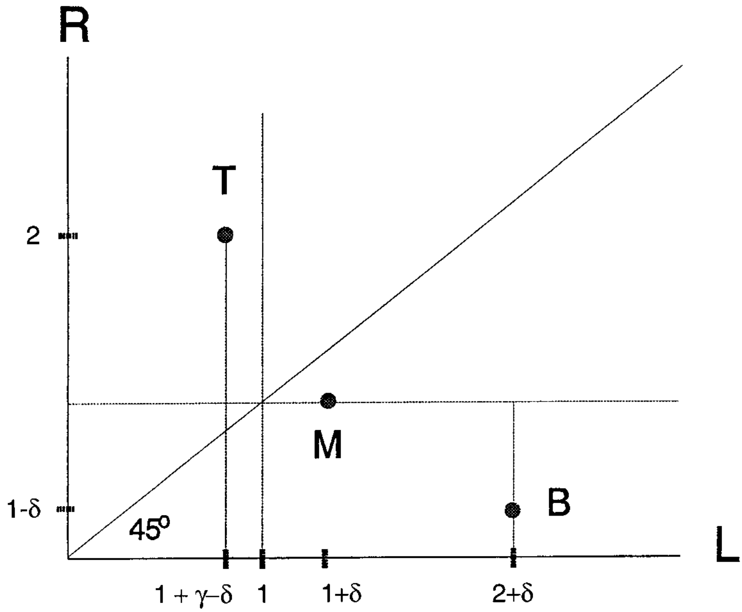

FIGURE 2

For the multiple-priors model, in this case where the state space consists of two elements, indifference curves are linear on either side of the certainty line but can be kinked there (subject to quasiconcavity). Therefore, M can be a best response only if the straight line through M and B intersects the certainty line to the right of the vertical line at $(1+\gamma-\delta)$; and this is true if and only if $\gamma \leqslant \delta + \delta^2$. Note that there is a sound economic argument at a descriptive level, (and arguably even at a normative level), for relaxing the assumption of probabilistic beliefs and thus admitting 'kinked' indifference curves. It is at least plausible that i be 'vague' or uncertain about j's choice of L or R, in which case her beliefs cannot be represented by a single probability measure. A noteworthy feature of the result concerning when M can be a best response is that the answer depends on γ. This is due to the quasiconcavity of utility that imposes roughly that the larger is γ, the less valuable are the incremental utility benefits of size δ, mentioned above, in playing M; and therefore the less likely is it that M can be a best response.

Finally, the ε-contamination model is a specialization of the multiple-priors model obtained by restricting the size of the kink that indifference curves can have along the certaintly line. From (4.7), it follows that the ratio of the slope in the upper cone to that in the lower cone can be no greater than $1+\{\varepsilon(1-\varepsilon)/[(1-\varepsilon)^2\, p^*(L)\, p^*(R)]\}$. The conditions under which M can be a best response are therefore readily derived.

5. EQUILIBRIUM

In the standard expected utility framework, Bandenburger and Dekel [10] have shown that (correlated) rationalizability is intimately related to an equilibrium concept, called *a posteriori* equilibrium, refining subjective correlated equilibrium. This section provides a parallel equilibrium recasting of $\mathscr{P}^*$-rationalizability in terms of what is called $\mathscr{P}^*$-*a posteriori* equilibrium, abbreviated $\mathscr{P}^*$-equilibrium. This is easily done by adapting the formulations and arguments in [10]. Nevertheless, given the decision-theoretic foundations provided in this paper for $\mathscr{P}^*$-equilibrium, I take the space to define it precisely and to describe its relation to $\mathscr{P}^*$-rationalizability.

Recall the normal form game (A_i, A_j, r_i, r_j) of Section 3, and the assumptions specified there. For what follows, it suffices that the model of preference $\mathscr{P}^*$ satsify conditions PREF1–PREF3 on the domain of all finite state spaces. Fix such a model of preference.

Define a $\mathscr{P}^*$-*a posteriori* equilibrium as a tuple $(\Omega, (\mathscr{H}_k, U_k, \sigma_k)_{k=i,j})$ where

- Ω is a finite state space
- $\mathscr{H}_k$ is an information partition for each player k
- $U_k \colon \Omega \to \mathscr{P}^*_k(\Omega)$ is k's conditional utility function, satisfying: $U_k(\omega, \cdot)$ knows $\mathscr{H}_k(\omega)$, the component of $\mathscr{H}_k$ containing ω
- $\sigma_k \colon (\Omega, \mathscr{H}_k) \to A_k$ is player k's measurable strategy function

and where for all $\omega \in \Omega$,

$$U_i(\omega; r_i(\sigma_i(\omega), \sigma_j(\cdot))) \geqslant U_i(\omega; r_i(a_i, \sigma_j(\cdot))) \qquad \forall a_i \in A_i, \tag{5.1}$$

and similarly for j. (Note that $r_i(\sigma_i(\omega), \sigma_j(\cdot))$ defines an act over Ω and therefore lies in the domain of each $U_i(\omega; \cdot)$. Similarly for the expression on the right.)

When $\mathscr{P}^*$ is the expected utility model of preference (for fixed vNM indices), this equilibrium is equivalent to that defined in [10]. That the equilibrium concept defined here is an attractive extension of the expected utility-based notion is confirmed by the equivalence with $\mathscr{P}^*$-rationalizability and by the foundations to follow in the next section.

The central result in this section is the following extension of [10, Proposition 2.1]:[8]

[8] As in [10, Proposition 2.1], this theorem could be stated in terms of the equivalence of utility payoffs.

THEOREM 5.1. *The profile* (a_i^*, a_j^*) *is* $\mathscr{P}^*$*-rationalizable* if and only if *there exists a* $\mathscr{P}^*$*-equilibrium* $(\Omega, (\mathscr{H}_k, U_k, \sigma_k)_{k=i,j})$ *and* $\omega^* \in \Omega$ *such that* $(a_i^*, a_j^*) = (\sigma_i(\omega^*), \sigma_j(\omega^*))$.

Proof. Only if: Define $\Omega \equiv R_i \times R_j$, $\sigma_i(a_i, a_j) \equiv a_i$, $\sigma_j(a_i, a_j) \equiv a_j$ and let $\mathscr{H}_k$ be the partition associated with knowledge of the kth coordinate of any $\omega = (a_i, a_j)$. It remains to define U_i (U_j may be defined similarly). Each $a_i \in R_i$ is a b.r. to some $u_i \in \mathscr{P}_i^*(A_j \mid R_j)$. Use u_i to define U_i as follows: For each $\omega = (a_i, a_j)$ and $f \in \mathscr{F}(R_i \times R_j)$, define

$$U_i(a_i, a_j; f) \equiv u_i(\hat{f}(a_i, \cdot)),$$

where $\hat{f}(a_i, \cdot)$ is the act over A_j satisfying $\hat{f}(a_i, \cdot) = f(a_i, \cdot)$ on R_j and $= r_i(a_i, \cdot)$ on $A_j \backslash R_j$. It follows from PREF2–3 that $U_i(a_i, a_j; \cdot)$ lies in $\mathscr{P}_i^*(\Omega)$. In addition, $U_i(a_i, a_j; \cdot)$ knows $\mathscr{H}_i(a_i, a_j) = \{a_i\} \times R_j$ because u_i knows R_j. The equilibrium condition (5.1) is satisfied because $U_i(a_i, a_j; a_i, \sigma_j(\cdot)) = u_i(r_i(a_i, \cdot)) \geqslant u_i(r_i(a_i', \cdot)) = U_i(a_i, a_j; a_i', \sigma_j(\cdot))$ $\forall a_i'$ in A_i. Finally, $\sigma_i(a_i^*, a_j^*) = a_i^*$ and similar assertions apply to j.

If: Let (a_i^*, a_j^*) be generated by a $\mathscr{P}^*$-equilibrium as described. Define $A_k^* \equiv \{\sigma_k(\omega) \colon \omega \in \Omega\}$ and show that $A_i^* \times A_j^* \subset R_i \times R_j$. For this it suffices to prove that

$$\text{each } a_i' \in A_i^* \text{ is a b.r. to some } u_i \in \mathscr{P}_i^*(A_j \mid A_j^*), \tag{5.2}$$

and similarly for j. Let $a_i' \in A_i^*$ and $\sigma_i(\omega') = a_i'$. Define u_i by

$$u_i(f) \equiv U_i(\omega'; f \circ \sigma_j(\cdot)), \qquad f \in \mathscr{F}(A_j).$$

PREF2 implies that $u_i \in \mathscr{P}_i^*(A_j)$. Moreover, u_i knows A_j^* because $f = g$ on $A_j^* \Rightarrow f \circ \sigma_j(\cdot) = g \circ \sigma_j(\cdot) \Rightarrow u_i(f) = u_i(g)$. Further, $u_i(r_i(a_i', \cdot)) \equiv U_i(\omega'; r_i(a_i', \sigma_j(\cdot))) \geqslant U_i(\omega'; r_i(a_i, \sigma_j(\cdot))) \equiv u_i(r_i(a_i, \cdot))$ for all $a_i \in A_i$, proving (5.2). ∎

6. FOUNDATIONS

The decision-theoretic foundation for $\mathscr{P}^*$-rationalizability is provided here. It takes the form of specifying the form of individual rationality and the knowledge of such rationality that characterize the selection of $\mathscr{P}^*$-rationalizable strategy profiles. In the special case of the expected utility model of preference $\mathscr{P}^* = \mathscr{P}^{EU}$, the familiar characterization [28, Theorems 5.2–5.3] of expected-utility-based rationalizability is obtained. More generally, for any model of preference $\mathscr{P}^*$, it is shown that

$\mathscr{P}^*$-rationalizability is characterized by 'rationality' and 'common knowledge of rationality', where each of these terms is defined appropriately in terms of the model of preference $\mathscr{P}^*$. Consequently, justification is provided for the procedure of iterated deletion of dominated strategies corresponding to any one of the dominance notions described in the previous section.

To achieve the desired characterization requires consideration of an extended state space for each agent, representing not only uncertainty about the opponent's choice of strategy but also about her 'type' that includes a description of her knowledge, beliefs or preferences. Use $\mathscr{P}(\cdot)$ to describe the exhaustive uncertainty facing each player.

To begin with, each player is uncertain about the strategy chosen by her opponent. Above we assumed that this strategic uncertainty was exhaustive. But in fact it is not, because i is uncertain also about j's preferences over A and these are relevant because knowledge of them would allow i to infer j's choice of strategy. Thus i's 'second-order state space' is $S_1 = A \times \mathscr{P}(A)$. Were this to represent all the uncertainty facing i, then we could identify each a_i with an act in $\mathscr{F}(A \times \mathscr{P}(A))$ and derive her strategy choice from her 'second-order preferences', an element in $\mathscr{P}(A \times \mathscr{P}(A))$. Similarly for j. But since i's second-order preferences are unknown to j and since they are useful for predicting what i will do, j faces the uncertainty represented by the state space $A \times \mathscr{P}(A) \times \mathscr{P}(A \times \mathscr{P}(A))$. Proceeding, one is led to the sequence of state spaces

$$S_0 = A, \qquad S_n = S_{n-1} \times \mathscr{P}(S_{n-1}), \qquad n \geqslant 0. \tag{6.1}$$

Each state space S_n is an *incomplete* description of the uncertainty facing i (or j) since given that S_n describes some of the uncertainty facing j, then i, in predicting j's behavior, faces uncertainty also about j's preferences over $\mathscr{F}(S_n)$.

For the above hierarchy to be well defined, it is necessary that each $\mathscr{P}(S)$ admit a topology such that $\mathscr{P}(\cdot)$ be compact-Hausdorff-valued. A further desideratum is that the infinite hierarchy represent, in a natural way, the *exhaustive* uncertainty facing each player. The key contribution of [14] is to show that both desiderata are achieved by $\mathscr{P}(\cdot)$ defined in Appendix A. In particular, the following construction of types spaces, extending [21] and [11], is valid.

THEOREM 6.1. *Define the correspondence $\mathscr{P}(\cdot)$ as in Appendix A. For any compact Hausdorff space A, there exists $T \subset \prod_0^\infty \mathscr{P}(S_n)$, such that when endowed with the induced product topology, T is compact Hausdorff and*

$$T \sim_{hmeo} \mathscr{P}(A \times T). \tag{6.2}$$

Denote by ψ the homeomorphism in (6.2). As above, each preference ordering can be identified with a unique utility function and so interpret $\psi(t_i)$ as a utility function for each $t_i \in T$.

To facilitate interpretation of the theorem, write $T_i = T_j = T$ and refer to these as spaces of 'types' for each player. Write the homeomorphism in the form

$$T_i \sim \mathscr{P}(A_j \times T_j) \qquad \text{and} \qquad T_j \sim \mathscr{P}(A_i \times T_i). \tag{6.3}$$

From the perspective of i, the state space $A_j \times T_j$ represents uncertainty about j's strategy and type. As above, i need also be concerned with j's preference ordering over $\mathscr{F}(A_i \times T_i)$, but, by the second homeomorphism in (6.3), this uncertainty is already represented by T_j. It follows that $A_j \times T_j$ is a complete or exhaustive state space for i and that a player may be described by and identified with her type.

It is now possible to provide a formal definition of rationality in the game context. Say that $(a_i, t_i) \in A_i \times T_i$ is *rational* if for all a_i' in A_i,

$$\psi(t_i)(r_i(a_i, \cdot)) \geqslant \psi(t_i)(r_i(a_i', \cdot)). \tag{6.4}$$

Each $r_i(a_i', \cdot)$ is an act over A_j. Therefore, it can be identified with an act over $A_j \times T_j$ and thus with an element in the domain of the preference ordering $\psi(t_i)$. Consequently, the utility maximizing nature of a_i is meaningfully expressed and provides natural meaning for rationality. Occasionally, the abbreviation "j is rational" will be used in lieu of "(a_j, t_j) is rational." Denote by Q_i the set of rational pairs in $A_i \times T_i$, and similarly for j.

This notion of rationality is weak in that $\mathscr{P}$ imposes weak restrictions on preferences. Consider a stronger definition of rationality that is tied to a model of preference $\mathscr{P}^*$. Say that i is $\mathscr{P}^*$*-rational* if (a_i, t_i) is rational (that is, lies in Q_i) and if i conforms to the model $\mathscr{P}_i^*$, which is naturally formalized by $\psi(t_i) \in \mathscr{P}_i^*(A \times T)$.

Consider now the hypothesis that both players are $\mathscr{P}^*$-rational and that this is common knowledge. Because types provide complete descriptions of players, this assumption must take formal expression through specification of subspaces $T_i^*, T_j^* \subset T$. These may be constructed as follows: Consider the sequence of sets $K_i^0 = K_j^0 = T$, and for $n \geqslant 1$,

$$K_i^n = \{t_i \in T_i : \psi(t_i) \in \mathscr{P}_i^*(A_j \times T_j \mid Q_j \cap [A_j \times K_j^{n-1}])\}; \tag{6.5}$$

and similarly for K_j^n. A pair $(t_i, t_j) \in K_i^n \times K_j^n$ indicates that i conforms to the model of preference $\mathscr{P}_i^*$, i knows that j is rational and that j conforms to the model $\mathscr{P}_j^*$, and so on to the nth order. Therefore, the natural candidates for T_i^* and T_j^* are

$$T_i^* = \bigcap_{n=0}^{\infty} K_i^n, \qquad T_j^* = \bigcap_{n=0}^{\infty} K_j^n. \tag{6.6}$$

In terms of these subspaces, the restrictions

$$(a_i, t_i) \in Q_i \cap [A_i \times T_i^*] \qquad \text{and} \qquad (a_j, t_j) \in Q_j \cap [A_j \times T_j^*] \tag{6.7}$$

formalize the assumptions that both players are rational, they conform to the model of preference $\mathscr{P}^*$, and that these facts are common knowledge. Moreover, 'knowledge' and preference (as represented by $\mathscr{P}^*$) are defined in a consistent fashion.

An outstanding question is whether T_i^* and T_j^* are nonempty. The following theorem, adapted from [14], confirms that they are nonempty and that they satisfy another appealing condition:

THEOREM 6.2. *Let $\mathscr{P}^*$ be a model of preference and adopt the definitions* (6.5)–(6.6). *Then $K_i^n \searrow_n T_i^*$ and $K_j^n \searrow_n T_j^*$, where these subspaces of types are nonempty and satisfy*

$$T_i^* \sim_\psi \mathscr{P}_i^*(A \times T \mid Q_j \cap [A \times T_j^*]), T_j^* \sim_\psi \mathscr{P}_j^*(A \times T \mid Q_i \cap [A \times T_i^*]).$$

This pair of homeomorphisms shows that the types subspaces T_i^* and T_j^* are 'closed' in the sense (suitably generalized) of 'beliefs closed' subspaces as defined in the Bayesian analysis [21]. A consequence is that $(t_i, t_j) \in T_i^* \times T_j^*$ formally models not only that $\mathscr{P}^*$-rationality is common knowledge, but also that this common knowledge is itself common knowledge, and so on to all orders. This confirms that our modeling of common knowledge is 'consistent' with the information-theoretic definition based on partitions or σ-algebras [2].[9]

Finally, turn to the objective of this section, namely to determine the implications of $\mathscr{P}^*$-rationality and common knowledge thereof. In terms of the preceding formal structure, this amounts to characterizing the strategy pairs (a_i, a_j) that are consistent with (6.7). Theorem 6.3 below provides this characterization, generalizing [28, Theorems 5.2–5.3] for the Bayesian setting.

THEOREM 6.3. *Let $\mathscr{P}^*$ be a model of preference. Then the strategy profile* (a_i^*, a_j^*) *is $\mathscr{P}^*$-rationalizable* if and only if *both players are $\mathscr{P}^*$-rational and this is common knowledge* (*where these assumptions are formalized in* (6.7)).

[9] See [11] for the relation between the information-theoretic and probabilistic approaches to defining common knowledge.

Proof. If: If i's type $t_i \in T_i^*$, then $t_i \in K_i^n \; \forall n$. It suffices to show that (recalling (6.5)),

$$t_i \in K_i^n \Rightarrow \psi(t_i) \text{ knows } R_j^n \times K_j^{n-1}. \tag{6.8}$$

For $n=1$, $\psi(t_i) \in \mathscr{P}_i^*(A_j \times T_j \mid Q_j)$ and $Q_j \subset R_j^1 \times T_j$. Assume for $n-1$ and prove for n. The assumption for t_i implies that $t_i \in K_i^{n-1}$ and that $\psi(t_i) \in \mathscr{P}_i^*(A_j \times T_j \mid A_j \times K_j^{n-1})$. The former implies that $\psi(t_i)$ knows Q_j, and so, by the conjunctive property of knowledge, $\psi(t_i) \in \mathscr{P}_i^*(A_j \times T_j \mid Q_j \cap [A_j \times K_j^{n-1}])$. By the induction hypothesis for j, $Q_j \cap [A_j \times K_j^{n-1}] \subset \{a_j : a_j$ is b.r. to some $u_j \in \mathscr{P}_j^*(A_i \mid R_i^{n-1})\} \times K_j^{n-1} = R_j^n \times K_j^{n-1}$. This proves (6.8).

It follows that $(a_i^*, t_i) \in Q_i \cap [A_i \times K_i^n] \Rightarrow a_i^*$ is a b.r. to $u_i \equiv mrg_{\mathscr{F}(A_j)}\, \psi(t_i)$ for some $\psi(t_i) \in \mathscr{P}_i^*(A_j \times T_j \mid R_j^n \times K_j^{n-1})$. But the latter implies, by the definition of marginals and PREF4, that $u_i \in \mathscr{P}_i^*(A_j \mid R_j^{n-1})$.[10] Hence, $a_i^* \in R_i^n$. This completes the proof of (a), because the preceding applies to every n.

Only if: Let (a_i^*, a_j^*) be rationalizable. It is enough to focus on player i. By hypothesis a_i^* is a b.r. to some $u_i^0 \in \mathscr{P}_i^*(A_j \mid R_j)$. We have to construct a type t_i satisfying (6.7). It is enough to show that for all $n \geqslant 1$,

$$\text{if } a_i^* \in R_i^n, \text{ then } \exists t_i \in T_i \text{ s.t. } (a_i^*, t_i) \in Q_i \cap [A_i \times K_i^{n-1}]. \tag{6.9}$$

Prove (6.9) inductively. First note that for any type $t_i = (u_i^n)_{n=0}^{\infty}$,

$$(a_i, t_i) \in Q_i \Leftrightarrow (a_i, mrg_{\mathscr{F}(A_j)}\, \psi(t_i)) \in Q_i^0 \equiv \{(a_i, u_i^0) : a_i \text{ is a b.r. to } u_i^0\}. \tag{6.10}$$

Take $n=1$. By hypothesis $\exists u_i^0 \in \mathscr{P}_i^*(A_j)$ such that a_i^* is a b.r. to u_i^0. The challenge is to 'extend' u_i^0 to a type, that is, to construct a type t_i such that $t_i = \psi(u)$ for some $u \in \mathscr{P}_i^*(A_j \times T_j)$ satisfying $mrg_{\mathscr{F}(A_j)}\, u = u_i^0$. It then follows from (6.10) that $(a_i^*, t_i) \in Q_i$ and hence that (6.9) holds. The existence of a suitable u is implied by (2.2). Alternatively, fix $t_j \in T_j$ and define u on $\mathscr{F}(A_j \times T_j)$ by $u(f) \equiv u_i^0(f(\cdot, t_j))$. Then $u \in \mathscr{P}_i^*(A_j \times T_j)$ by PREF2 and its marginal over A_j equals u_i^0.

Assume (6.9) for $n-1$ and prove for n. By hypothesis a_i^* is a b.r. to $u_i^0 \in \mathscr{P}_i^*(A_j \mid R_j^{n-1})$. The induction hypothesis implies that for every $a_j \in R_j^{n-1}$ there exists $t_j[a_j]$ satisfying: for every $a_j \in R_j^{n-1}$,

$$(a_j, t_j[a_j]) \in Q_j \cap (A_j \times K_j^{n-2}). \tag{6.11}$$

[10] We are using: (i) If $\psi(t)$ knows a closed subset $C \subset A \times T$, then $mrg_{\mathscr{F}(A)}\, \psi(t) \in \mathscr{P}(A)$ knows $proj_A\, C$; and (ii) the latter projection is closed, because of the compactness of A and T.

Define $t_j[a_j] \in T_j$ arbitrarily for $a_j \in A_j \backslash R_j^{n-1}$ and define u on $\mathscr{F}(A_j \times T_j)$ by

$$u(f) \equiv u_i^0(f(\cdot, t_j[\cdot])).$$

Then PREF2, (6.11) and the hypothesis that u_i^0 knows R_j^{n-1} imply that $u \in \mathscr{P}_i^*(A_j \times T_j \mid Q_j \cap (A_j \times K_j^{n-2}))$. It follows that $t_i = \psi^{-1}(u)$ satisfies (6.9). ∎

This theorem applies to each of the specific models of preference described in Section 4. It is necessary only to indicate how the description of each of those models of preference is to be completed so as to include infinite state spaces like $A \times T$. Roughly, this is done simply by replacing A by $A \times T$ throughout the preceding definitions, taking care to add the suitable technical details. For example, the expected utility, ordinal expected utility and probabilistically sophisticated models make use of probability measures that are now taken to be Borel regular probability measures on $A \times T$; the set of such measures is denoted $M(A \times T)$ and replaces $M(A)$ in the definitions. In the multiple-priors model, in conformity with [15], the set of priors Δ is taken to be a weak*-closed and convex set of finitely additive Borel probability measures on $A \times T$. Similarly for the set of priors C appearing in the ε-contamination model.

7. CONCLUDING REMARKS

This paper has extended theorems in [28] and [10] regarding the implications for finite normal form games of the hypothesis of rationality and common knowledge or rationality. The extension took the form of generalizing the definition of rationality beyond subjective expected utility maximization. Some concrete examples of ‘admissible’ forms of rationality were provided, including probabilistic sophistication, conformity with the multiple-priors model and monotonic preferences. Other examples may occur to the reader. Another prominent ‘nonexpected utility’ model of preference that has been proposed in order to model uncertainty aversion is Choquet expected utility [26]. As formulated in [14], Choquet expected utility is a model of preference in the formal sense of this paper, namely it satisfies conditions PREF1–4. Therefore, it lies within the scope of the theorems of this paper.

The extension provided here is useful for providing perspective for and deeper understanding of the expected utility-based theorems. It is hoped that it will help also to lay the groundwork for future uses of ‘nonexpected utility’ preferences in applied game theoretic modeling, for example, in order to explore the intuitively plausible hypothesis that aversion to uncertainty or ambiguity may be important in strategic situations. Alternative foundations for future applications may be found in [13, 16, 17 and 22], who provide equilibrium concepts for normal form games in which players' preferences are in the multiple-priors or Choquet expected utility classes.

Extensions of this paper may also serve as areas for future research. While results concerning $\mathscr{P}^*$-rationalizability and $\mathscr{P}^*$-*a posteriori* equilibrium apply for general models of preference $\mathscr{P}^*$, our derivation of the corresponding 'dominance' notion was more limited. For example, our arguments based on Theorems of the Alternative do not seem to apply to the Choquet expected utility model. The extension to 'independent' rationalizability for n players also remains to be done. Finally, investigation of foundations for alternative equilibrium concepts seems worthwhile. Lo [17] provides epistemic conditions characterizing his equilibrium notion.

APPENDIX A

For any compact Hausdorff S, $\mathscr{F}(S)$ denotes the set of acts over S, that is, the set of all Borel measurable functions on S that have values in the compact interval X. We describe $\mathscr{P}(S)$, the class of *regular* preferences over $\mathscr{F}(S)$. See [14] for more detailed description and interpretation.

First designate various subsets of $\mathscr{F}(S)$. Call an act *simple* if its range is finite. Call an act f upper semicontinuous (usc) if all sets of the form $\{s: f(s) \geqslant \kappa$ are closed. Similarly, f is lower semicontinuous (lsc) if all sets of the form $\{s: f(s) > \kappa\}$ are open. Denote by $\mathscr{F}^u(S)$ and $\mathscr{F}^l(S)$ the sets of simple usc and simple lsc acts respectively. The outcome $x \in X$ also denotes the corresponding constant act.

Define $\mathscr{P}(S)$ as the set of all utility functions u: $\mathscr{F}(S) \to X$ satisfying:

U.1. Certainty Equivalence: $u(x) \equiv x$.

U.2. Weak monotonicity: $f' \geqslant f \Rightarrow u(f') \geqslant u(f)$.

U.3. Inner Regularity: $u(f) = \sup\{u(g): g \leqslant f, g \in \mathscr{F}^u(S)\}$, $\forall f \in \mathscr{F}(S)$.

U.4. Outer Regularity: $u(g) = \inf\{u(h): h \geqslant g, h \in \mathscr{F}^l(S)\}$, $\forall g \in \mathscr{F}^u(S)$.

It is shown in [14] that there is a one-to-one correspondence between this class of utility functions and a suitably specified class of preference orderings. This justifies our referring to elements of $\mathscr{P}(S)$ interchangeably as utility functions or preference orderings. The regularity conditions (U.3) and (U.4) 'mimic' the property of regularity for probability measures; think of u as a measure and replace g, h and f by closed, open and measurable subsets of S, respectively, noting that the characteristic (or indicator) function for a closed set is usc, and so on. If S is finite as in Sections 3–5, then all acts are both usc and lsc and so conditions (U.3) and (U.4) are trivially satisfied.

The topology τ on $\mathscr{P}(S)$ is that generated by the subbasis consisting of sets of the form

$$\{u\colon u(g) < \kappa\} \qquad \{u\colon u(h) > \kappa\}, \tag{A.1}$$

where κ varies over the reals and g and h vary over $\mathscr{F}^u(S)$ and $\mathscr{F}^l(S)$, respectively. That is, τ is the coarsest topology on $\mathscr{P}(S)$ that makes the mapping $u \mapsto u(f)$ usc for every $f \in \mathscr{F}^u(S)$ and lsc for every $f \in \mathscr{F}^l(S)$. Using the above identification between usc or lsc acts and closed or open sets, there is a formal similarity to the weak convergence topology on the set of Borel measures. More importantly, the topology τ makes $\mathscr{P}(S)$ compact Hausdorff if S is compact Hausdorff.

APPENDIX B

Here prove that the negation of (4.4) is necessary and sufficient for $a^* \in A_i$ to be player i's b.r. for some set of priors Δ over A_j. For convenience, denote $A_i \backslash \{a^*\}$ by A_i^* or simply A^*. In terms of the notation in the text, the best response property requires that Δ satisfy

$$\min_{m \in \Delta} \int g_{a^*}\, dm \geqslant \min_{m \in \Delta} \int g_a\, dm \qquad \forall a \in A^*. \tag{B.1}$$

Denote by m_a a measure where the minimum on the right is attained. Then we must have

$$\int g_{a^*}\, dm_{a'} \geqslant \int g_a\, dm_a \qquad \forall a, a' \in A^*. \tag{B.2}$$

In fact, the existence of $\{m_a\}_{a \in A^*}$ is also sufficient for (B.1), because we could define Δ to be the convex hull of $\{m_a\}_{a \in A^*}$. Therefore, we proceed to characterize the tuples of acts $(g_{a^*}, \{g_a\}_{a \in A^*})$ for which there exist probability measures $\{m_a\}_{a \in A^*}$ solving (B.2). For vector inequalities $x \geqslant y$ indicates weak inequality for all components, while $x > y$ indicates that in addition $x \neq y$.

Supplement the above inequalities with the non-negativity condition for probabilities $m_a > 0$. Then the normalizations $\sum_{a_j} m_a(a_j) = 1$ for all $a \in A^*$ can be replaced by the equalities $\sum_{a_j} m_a(a_j) = \sum_{a_j} m_{a'}(a_j)$ $\forall a, a' \in A^*$. The advantage of such a reformulation is that now the supplemented system of inequalities is a homogeneous system of linear inequalities in $\{m_a\}_{a \in A^*}$, a vector in some Euclidean space. More precisely, the complete set of restrictions on $\{m_a\}_{a \in A^*}$ can be expressed in the form

$$x > 0, \qquad Bx \geqslant 0 \text{ and } Dx = 0, \tag{B.3}$$

for suitable matrices B and D, where x is the vector of dimension $|A^*| \times |A_j|$ obtained by stacking the measures $\{m_a\}_{a \in A^*}$. Tucker's Theorem of the Alternative [20, p. 29] characterizes the conditions under which such systems admit solutions. After some tedious but elementary algebraic manipulations, we conclude: There does *not* exist a solution to our system if and only if $\exists \{\alpha_a, \beta_a\}_{a \in A^*} \subset [0, 1]$, $\sum \alpha_a = \sum \beta_a = 1$, $\{\gamma_a\}_{a \in A^*}$, $\sum \gamma_a = 0$ and a mapping $h_a \colon A_j \to (0, \infty)$, for each $a \in A^*$, such that

$$\alpha_a g_a(a_j) - \beta_a g_{a^*}(a_j) = h_a(a_j) + \gamma_a, \qquad a_j \in A_j, a \in A^*.$$

We can eliminate $\{h_a, \gamma_a\}$ by expressing the preceding condition in the equivalent forms

$$\min_{a_j} [\alpha_a g_a(a_j) - \beta_a g_{a^*}(a_j)] > \gamma_a, \text{ or}$$

$$\sum_{a \in A^*} \min_{a_j} [\alpha_a g_a(a_j) - \beta_a g_{a^*}(a_j)] > 0.$$

This proves the assertion regarding (4.4).

Turn to the proof of the assertion regarding the ε-dominance condition (4.8). To take account of the restricted nature of sets of priors $\Delta \in \Upsilon^\varepsilon$, one need only replace each m_a above by $(1-\varepsilon) p^* + \varepsilon p_a$ and search for p^* and $\{p_a\}_{a \in A^*}$ that satisfy (B.2) and appropriate non-negativity and summation conditions. This complete set of restrictions can again be expressed in the form (B.3), with redefined matrices B and D and with the vector x composed of p^* and $\{p_a\}_{a \in A^*}$. Tucker's Theorem delivers the desired result. ∎

REFERENCES

1. F. J. Anscombe and R. Aumann, A definition of subjective probability, *Ann. Math. Stat.* **34** (1963), 199–205.
2. R. Aumann, Agreeing to disagree, *Ann. Stat.* **4** (1976), 1236–1239.
3. R. Aumann, Correlated equilibrium as an expression of Bayesian rationality, *Econometrica* **55** (1987), 1–18.
4. P. Battigalli, Strategic rationality orderings and the best rationalization principle, mimeo, 1994.

5. D. Bernheim, Rationalizable strategic behavior, *Econometrica* **52** (1984), 1007–1028.
6. T. Bewley, "Knightian Decision Theory: Part I," Cowles Foundation #807, 1986.
7. L. Blume, A. Brandenburger, and E. Dekel, Lexicographic equilibrium and equilibrium refinements, *Econometrica* **59** (1991), 81–98.
8. T. Borgers, Pure strategy dominance, *Econometrica* **61** (1993), 423–430.
9. A. Brandenburger, Knowledge and equilibrium in games, *J. Econ. Perspectives* **6** (1992), 83–101.
10. A. Brandenburger and E. Dekel, Rationalizability and correlated equilibria, *Econometrica* **55** (1987), 1391–1402.
11. A. Brandenburger and E. Dekel, Hierarchies of beliefs and common knowledge, *J. Econ. Theory* **59** (1993), 189–198.
12. E. Dekel, Z. Safra, and U. Segal, Existence and dynamic consistency of Nash equilibrium with non-expected utility preferences, *J. Econ. Theory* **55** (1991), 229–246.
13. J. Dow and S. R. C. Werlang, Nash equilibrium under uncertainty: Breaking down backward induction, *J. Econ. Theory* **64** (1994), 305–324.
14. L. G. Epstein and T. Wang, 'Beliefs about beliefs' without probabilities, *Econometrica* **64** (1996), 1343–1373.
15. I. Gilboa and D. Schmeidler, Maxmin expected utility with nonunique prior, *J. Math. Econ.* **18** (1989), 141–153.
16. P. Klibanoff, Uncertainty, decision and normal form games, mimeo, 1994.
17. K. C. Lo, Equilibrium in beliefs under uncertainty, *J. Econ. Theory*, forthcoming.
18. K. C. Lo, "Nash Equilibrium without Mutual Knowledge of Rationality," mimeo, 1995.
19. M. Machina and D. Schmeidler, A more robust definition of subjective probability, *Econometrica* **60** (1992), 745–780.
20. O. Mangasarian, "Nonlinear Programming," McGraw–Hill, New York, 1969.
21. J. F. Mertens and S. Zamir, Formulation of Bayesian analysis for games with incomplete information, *Int. J. Game Theory* **14** (1985), 1–29.
22. S. Mukerji, A theory of play for games in strategic form when rationality is not common knowledge, mimeo, 1995.
23. D. Pearce, Rationalizable strategic behavior and the problem of perfection, *Econometrica* **52** (1984), 1029–1050.
24. A. Rubinstein, Comments on the interpretation of game theory, *Econometrica* **59** (1991), 909–924.
25. L. Savage, "The Foundations of Statistics," Wiley, New York, 1954.
26. D. Schmeidler, Subjective probability and expected utility without additivity, *Econometrica* **57** (1989), 571–587.
27. D. Stahl, Lexicographic rationality, common knowledge and iterated admissibility, mimeo, 1991.
28. T. C. C. Tan and S. R. Werlang, The Bayesian foundations of solution concepts of games, *J. Econ. Theory* **45** (1988), 370–391.