Modelling Beliefs in Games with Generalized Preferences

Rupa Athreya*

Department of Economics
Yale University

This version: May 2001

1 Introduction

Traditional decision theory treats risk (situations where probabilities of events are known) as formally equivalent to uncertainty (situations in which the probabilities of events are unknown). The subjective expected utility (SEU) model axiomatized by Savage (1954) has been the most important theory in analysing decision making under uncertainty. Expected utility (EU) has been the accepted basis for analysing decision making in games as well. Using the EU model to represent players' preferences, a number of solution concepts have been developed, most prominent among them being Nash Equilibrium.

However, the descriptive validity of the SEU model has been questioned. In recent years generalizations of, and alternatives to, the SEU framework have been developed, where the decision maker does not have point beliefs that can be represented as in Savage. A growing literature has attempted to extend this theory to interactive situations[1]. Players in games are modelled as facing uncertainty regarding opponents' strategies and the standard solution concepts are generalized to account for deviations from subjective expected utility maximization.

Savage (1954) shows how properties of a decision maker's probabilistic beliefs can be deduced from primitive consistency axioms on preferences. Different restrictions on conditional preferences generate different notions of qualitative belief. Under the SEU axioms, these are equivalent, so that the "subjective probability" generated by SEU preferences has a natural epistemic interpretation: belief can be identified with "full belief" or belief with probability one. However, when individuals are not expected utilily

*Many thanks to Stephen Morris and Ben Polak for valuable discussions, support and encouragement. I am also grateful to Larry Blume, Adam Brandenburger, Juan Dubra and Matthew Ryan for comments, and to participants at the Yale Summer Seminar and the Game Theory Reading Group for useful suggestions.
Address: 28 Hillhouse Avenue, New Haven, CT 06511. e-mail: rupa.athreya@yale.edu

[1] See, for example, Dow and Werlang (1994), Eichberger and Kelsey (1998), Epstein (1997), Klibanoff (1996), Lo (1996), Marinacci (1996), Mukerji (1997), Ryan (1999).

maximizers, the alternative notions of belief generated by assumptions on preferences are not necessarily equivalent. That is, qualitative notions of belief arising out of different restrictions on preferences vary in their implications. For example, when players' preferences belong to the maxmin expected utility (MEU) class, "full belief" can be thought of as corresponding to a situation where all beliefs in the agent's belief set place probability one on the event that is believed; weaker notions of belief might relax this notion to require, for instance, that beliefs only place some positive weight on the event that is believed, or that events which are not believed are assigned zero weight by at least one belief, but not necessarily by all of them.

When different notions of belief are no longer equivalent, the choice of which notion of belief to employ becomes critical in extending decision theory to games. Alternative representations of preferences, and notions of belief, will mean that restrictions on the rationality of opponents lead to different epistemic conditions for solution concepts.

Morris (1997) provides a general framework for preference-based belief which allows comparison of different notions of belief in order to determine which notions are 'stronger' or 'weaker' than others. He studies belief operators, which provide a semantic characterisation of belief by identifying, for each state of the world, the events that the decision maker believes. Belief operators defined from the decision maker's preferences can be characterised using standard preference representations, and provide epistemic models for decision making in both expected utility and non-expected utility environments.

The purpose of this paper is to bring together these developments in the decision-theoretic foundations of games by setting out a general framework for characterising dominance and equilibrium in normal form games with players whose preferences may deviate from the expected utility model. The standard framework considers players with SEU preferences; knowledge is defined as belief with probability one, and epistemic conditions for dominance and equilibrium are derived. In this paper, I relax the axioms on preferences, allowing them to deviate from expected utility, and investigate the implications of varying belief of rationality for dominance and equilibrium.

I first show that common belief of rationality, where belief is defined in the sense of Savage, and weak monotonicity assumptions on preferences, lead players to play pure strategies that are iteratively undominated by other pure strategies. However, when the notion of belief is weakened, then it is not always possible to rule out play of iteratively strictly dominated strategies. I also show that when preferences are admissible and beliefs respect strict dominance, common belief of rationality implies one round of deletion of weakly dominated strategies, with subsequent rounds of deletion confined to strictly dominated strategies.

In a closely related paper, Epstein (1997) considers rationalizability and equilibrium when players' preferences deviate from expected utility maximization. However, he retains Savage belief and investigates the implications of varying the notion of rationality, relating a general notion of rationalizability to the standard model of expected utility and to pure strategy dominance. The exercise carried out in this paper is to relate different notions of belief and to compare the implications of common belief of rationality, appropriately defined, across representations of preferences.

I also investigate epistemic conditions for pure-strategy Nash equilibrium, defined to be a pair of pure-strategy best responses. In the standard model, mutual belief of actions implies that actions form a Nash equilibrium (Aumann and Brandenburger (1995)); I show first that with general preferences, mutual Savage belief of actions is sufficient for actions to be a Nash equilibrium, but that when belief is weakened to "weak" belief, as defined in Morris (1997), there exists a class of preferences for which mutual belief of actions does not imply that players play a Nash equilibrium.

The paper is organized as follows. In section 2, I review briefly the decision-theoretic framework and recent extensions. In section 3, I extend this to two-player situations. I define an interactive belief system and prove some preliminary results regarding properties of interactive beliefs. Section 4 considers pure strategy dominance, and characterises conditions for iterative strict and weak dominance. Section 5 explores epistemic conditions for equilibrium with generalized preferences.

2 Preferences and Beliefs

2.1 Preference preliminaries

The traditional approach to modelling choice under uncertainty in economics and decision theory combines an exogenous framework for modelling belief with a set of preferences. However, a given set of preferences over acts with state-contingent outcomes, and beliefs over those states, will not in general be reconcilable without further restrictions. Hence a more natural approach in this context is that due to Savage (1954), where individual beliefs are derived from restrictions on preferences, so that beliefs can be thought of as being "revealed" by behavior rather than imposed *a priori* by the modeller.

I begin by outlining a framework for preferences, and then defining belief in those terms. Let Ω be a finite set of states of the world. Let the set of "acts" be the set of all functions from Ω to $\mathbf{R}$, the real line, so that acts can be thought of as vectors in $\mathbf{R}^\Omega$. Thus act $x \in \mathbf{R}^\Omega$ yields the prize x_ω in state $\omega \in \Omega$. For $E \subseteq \Omega$, x_E denotes the tuple $\{x_\omega\}_{\omega \in E}$. I denote by $-E$ the complement of E in Ω and use the notation $x_{\{\omega\}}$ or x_ω and $x_{-\{\omega\}}$ or $x_{-\omega}$ interchangeably. At each state of the world, ω, an individual is assumed to have a preference relation, $\succeq_\omega$, over acts, where $x \succeq_\omega y$ implies that, at ω, act x is at least as good as act y. Strict preference and indifference are defined in the usual fashion.

In the standard EU framework, an individual's beliefs are represented by a single countably additive probability measure on the set of states. Preference relations $\{\succeq_\omega\}_{\omega \in \Omega}$ on $\mathbf{R}^\Omega$ have an EU representation if there exists, for each state $\omega \in \Omega$, an increasing and continuous utility function $u_\omega : \mathbf{R} \to \mathbf{R}$ and a countably additive probability measure p_ω on Ω such that, for all $x, y \in \mathbf{R}^\Omega$,

$$x \succeq_\omega y \Leftrightarrow \sum_{\omega' \in \Omega} p_\omega(\omega') u_\omega(x_{\omega'}) \geq \sum_{\omega' \in \Omega} p_\omega(\omega') u_\omega(y_{\omega'}).$$

Two alternative representations of preferences which relax this restriction are the *maxmin expected utility* (MEU) framework and the *Choquet expected utility* (CEU) model. Preference relations $\{\succeq_\omega\}_{\omega \in \Omega}$ on $\mathbf{R}^\Omega$ have an MEU representation if there exists, for each state $\omega \in \Omega$, an increasing and continuous utility function $u_\omega : \mathbf{R} \to \mathbf{R}$ and a unique, nonempty, closed and convex conditional belief set C_ω of countably additive probability measures on Ω such that, for all $x, y \in \mathbf{R}^\Omega$,

$$x \succeq_\omega y \Leftrightarrow \min_{p_\omega \in C_\omega} \left\{ \sum_{\omega' \in \Omega} p_\omega(\omega') u_\omega(x_{\omega'}) \right\} \geq \min_{p_\omega \in C_\omega} \left\{ \sum_{\omega' \in \Omega} p_\omega(\omega') u_\omega(y_{\omega'}) \right\}.$$

In the case of CEU preferences, the probability measure is non-additive. A non-additive probability measure (sometimes known as a *capacity*) is a function $\nu : 2^\Omega \to \mathbf{R}_+$, with $\nu(\emptyset) = 0$, $\nu(\Omega) = 1$ and $\nu(E) \leq \nu(F)$ if $E \subseteq F$. The *range* of act x is the set of values attained on Ω. Let T be any finite ordered subset of the real line containing the range of x, so that $\{r \in \mathbf{R} | \, r = x_\omega \text{ for some } \omega \in \Omega\} \subseteq T = \{r_1, \cdots, r_K\} \subseteq \mathbf{R}$ and $r_1 > r_2 > \cdots > r_K$. Letting $r_{K+1} = 0$, the expected value of x is

$$\mathbf{E}_\nu(x) = \sum_{k=1}^{K} (r_k - r_{k+1}) \, \nu(\{\omega \in \Omega | x_\omega \geq r_k\}).$$

This definition reduces to the usual notion of expected value if ν is additive. Write $u_\omega(x)$ for the vector $\{u_\omega(x_\omega)\}_{\omega \in \Omega}$. Preference relations $\{\succeq_\omega\}_{\omega \in \Omega}$ on $\mathbf{R}^\Omega$ have a CEU representation if there exists, for each state $\omega \in \Omega$, an increasing and continuous utility function $u_\omega : \mathbf{R} \to \mathbf{R}$ and a non-additive probability measure on ν_ω such that, for all $x, y \in \mathbf{R}^\Omega$,

$$x \succeq_\omega y \Leftrightarrow \mathbf{E}_{\nu_\omega}(u_\omega(x)) \geq \mathbf{E}_{\nu_\omega}(u_\omega(y)).$$

Lexicographic preferences allow decision makers to take into account events which are assigned probability zero ex ante. Say that x is lexicographically greater than y $[x \geq_L y]$ if $y_i > x_i$ implies $x_h > y_h$ for some $h < i$. Preference relations $\{\succeq_\omega\}_{\omega \in \Omega}$ on $\mathbf{R}^\Omega$ have a lexicographic expected utility (LEU) representation if there exists, (1) a positive integer J, (2) for each state $\omega \in \Omega$, an increasing and continuous utility function $u_\omega : \mathbf{R} \to \mathbf{R}$ and (3) for each $\omega \in \Omega$ and $j = 1, \cdots, J$, a countably additive probability measure p^j_ω on Ω such that, for all $x, y \in \mathbf{R}^\Omega$,

$$x \succeq_\omega y \Leftrightarrow \left\{ \sum_{\omega' \in \Omega} p^j_\omega(\omega') u_\omega(x_{\omega'}) \right\}_{j=1}^{J} \geq_L \left\{ \sum_{\omega' \in \Omega} p^j_\omega(\omega') u_\omega(y_{\omega'}) \right\}_{j=1}^{J}.$$

In order to ensure that the preference relation is a non-trivial ordering, I shall impose the following restrictions: reflexivity [P1], non-triviality [P2], transitivity [P3] and completeness [P4]. In addition to these, there are various monotonocity assumptions that may be made. Say that $x \geq y$ if $x_\omega \geq y_\omega$ for all $\omega \in \Omega$; $x > y$ if $x \geq y$ and $x_\omega > y_\omega$ for some $\omega \in \Omega$; $x \gg y$ if $x_\omega > y_\omega$ for all $\omega \in \Omega$.

P5 : Preference relations are *admissible* if $x > y \Rightarrow x \succ_\omega y$ and $x \geq y \Rightarrow x \succeq_\omega y$, for all $\omega \in \Omega$.

A weaker assumption, which is implied by [P5], is the following:

P5* : Preference relations are *monotone* if $x \gg y \Rightarrow x \succ_\omega y$ and $x \geq y \Rightarrow x \succeq_\omega y$, for all $\omega \in \Omega$.

Two more substantive assumptions are:

P6 : Preference relations satisfy *non-null statewise monotonicity* if, for all $\omega, \omega' \in \Omega$, *either* $(x_{\omega'}, z_{-\omega'}) \sim_\omega (y_{\omega'}, z_{-\omega'})$ for all $x, y, z \in \mathbf{R}^\Omega$ *or* $(x_{\omega'}, z_{-\omega'}) \succ_\omega (y_{\omega'}, z_{-\omega'})$ for all $x \gg y, z$.

P7 : Preference relations are *continuous* if, for all for all $\omega \in \Omega$, the set $\{x \in \mathbf{R}^\Omega \mid x \succeq_\omega y\}$ is closed.

2.2 Beliefs, possibility and supports

Under certain conditions, results about belief on a finite state space can be represented by two equivalent formalisms: belief operators and possibility relations. A belief operator specifies, for each subset of the state space (or event), the set of states where the individual believes that event is true. A belief operator is thus a mapping $B : 2^\Omega \to 2^\Omega$, with the interpretation that the individual believes event $E \subseteq \Omega$ at state $\omega \in \Omega$ if and only if $\omega \in B(E)$.

A possibility correspondence specifies, for each state of the world, which states the individual thinks are possible: it is a mapping from the uncertainty space to its subsets, $P : \Omega \to 2^\Omega$. The interpretation is that if $\omega' \in P(\omega)$, the state ω' is thought possible when the true state is ω.

Definition 1 *The operator* $B : 2^\Omega \to 2^\Omega$ represents *the possibility relation* P *if*

$$B(E) = \{\omega \in \Omega | \, P(\omega) \subseteq E\}$$

$$P(\omega) = \bigcap \{F \subseteq \Omega | \, \omega \in B(F)\}$$

The operator B *is a* normal *belief operator if there exists* P *such that* B *represents* P.

We are now in a position to define belief in terms of preferences. Savage's notion of belief captures the idea that if the individual's preferences never depend on anything that happens when event E does not occur, then the individual believes E. If the individual is ever concerned about what happens when E does not occur, then the individual cannot believe E. This is closely related to Savage's notion of *null* events: an individual (Savage) believes an event E if the complement of E is Savage-null.

Definition 2 *Belief operator* B^{**} *represents* Savage belief *of preference relations* $\{\succeq_\omega\}_{\omega \in \Omega}$ *if*

$$B^{**}(E) = \{\omega \in \Omega | (x_E, z_{-E}) \succeq_\omega (x_E, y_{-E}) \ \forall x, y, z \in \mathbf{R}^\Omega\}.$$

The associated possibility correspondence is defined as:

$$P^{**}(\omega) = \{\omega' \in \Omega | \exists x, y, z \in \mathbf{R}^\Omega \text{ such that } (x_{\omega'}, z_{-\omega'}) \succ_\omega (y_{\omega'}, z_{-\omega'})\}.$$

This notion of belief arises naturally in the standard expected utility model, where an event is believed if and only if it is assigned probability one by the decision maker's (subjective) probability measure. It is also possible to define alternative notions of belief.

Definition 3 *Belief operator B^* represents* strong belief *of preference relations $\{\succeq_\omega\}_{\omega \in \Omega}$ if*

$B^*(E) = \{\omega \in \Omega | (x_E, z_{-E}) \succeq_\omega (y_E, v_{-E}) \ \ \forall x \gg y, x, y, v, z \in \mathbf{R}^\Omega\}$.

The associated possibility correspondence is defined as:

$P^*(\omega) = \{\omega' \in \Omega | \exists x \gg y \gg z \in \mathbf{R}^\Omega \text{ such that } (x_{\omega'}, z_{-\omega'}) \succ_\omega y\}$.

Strong belief formalizes the intuition that strict dominance on the event that is believed should be sufficient to determine the direction of preference, regardless of the outcome on the rest of the space. When preferences have an LEU representation, an event is strongly believed if it is assigned probability one by the decision maker's first-order beliefs. Savage belief, by contrast, requires that *each* of the decision maker's beliefs assign probability one to the event that is believed.

We may also consider the following strengthening of strong belief, which I will call strong* belief.

$$\tilde{B}(E) = \{\omega \in \Omega | (x_E, z_{-E}) \succ_\omega (y_E, v_{-E}) \ \ \forall x \gg y, x, y, v, z \in \mathbf{R}^\Omega\}.$$

The associated possibility correspondence then becomes:

$$\tilde{P}(\omega) = \{\omega' \in \Omega | \exists x \gg y \gg z \in \mathbf{R}^\Omega \text{ such that } (x_{\omega'}, z_{-\omega'}) \succeq_\omega y\}.$$

A notion of belief which utilizes a much stronger notion of possibility is weak belief.

Definition 4 *Belief operator B represents* weak belief *of preference relations $\{\succeq_\omega\}_{\omega \in \Omega}$ if*

$$B(E) = \{\omega \in \Omega | \ P(\omega) \subseteq E\}.$$

where

$$P(\omega) = \{\omega' \in \Omega | (\forall x \gg y)(\exists z \ll y) \text{ such that } (x_{\omega'}, z_{-\omega'}) \succ_\omega y\}.$$

When preferences have an MEU representation, a state is weakly possible if it is assigned positive probability by *each* of the probability measures in the decision maker's set of conditional beliefs. An event is weakly believed if it contains every state assigned positive probability by all the measures in the belief set. By contrast, Savage or strong belief would require that the event believed is assigned probability one by *all* measures in the set of beliefs, with a state being considered possible if there exists a measure in the belief set which assigns positive probability to it.

3 Interactive beliefs

I now extend this framework to situations with two agents. I consider 2-player, finite action games, consisting of a finite set of players $I = \{1, 2\}$, finite pure strategy spaces S_i $(i = 1, 2)$, with typical element s_i, and payoff functions $g_i : S_1 \times S_2 \to \mathbf{R}$. Following Aumann and Brandenburger (1995), I define an interactive belief system as follows. Each player has a finite set of types, T_i, and

- for each type t_i, $i \neq j$, a preference relation $\succeq_{t_i}$ over the set of "acts" $\mathbf{R}^{T_j}$, and
- an action function $f_i : T_i \to S_i$, which maps each type of a player to his strategy space.

A type is thus a formal description of a player's actions and preferences (and thus beliefs). For simplicity, I restrict payoff functions to be the same across all types of a given player, and assume that these are "common knowledge" (loosely speaking). The subjective uncertainty faced by player i is then represented by player j's type space. Thus, strategy choice s_i by player i induces the act $[g_i(s_i, f_j(t_j))]_{t_j \in T_j}$, which is a vector of (utility) outcomes.

I assume that each player knows his own type. This implies that from the point of view of type t_i of player i, while the state space consists of pairs $(t_i, t_j), t_j \in T_j$, the relevant uncertainty for this type of player i is summarized in the space T_j.

Definition 5 *A* belief operator *for player i maps subsets of the state space $\Omega = T_1 \times T_2$ to itself: i.e., $B_i : 2^\Omega \to 2^\Omega$. Consider events of the form $A = A_1 \times A_2$, where $A_1 \subseteq T_1$ and $A_2 \subseteq T_2$. Since each player knows his own type, we can define operators $\hat{B}_1$ and $\hat{B}_2$, where $\hat{B}_1 : 2^{T_2} \to 2^{T_1}$ and vice versa, such that $B_1(A_1 \times A_2) = (A_1 \cap \hat{B}_1(A_2)) \times T_2$ and $B_2(A_1 \times A_2) = T_1 \times (A_2 \cap \hat{B}_2(A_1))$.*

The belief operator $\hat{B}_1$ specifies the subsets of T_2 that player 1 believes: that is, $t_1 \in \hat{B}_1(A_2)$ if and only if type t_1 of player 1 believes that player 2 is some type in A_2. Then the event $B_1(A_1 \times A_2)$, which can be read as "player 1 believes the event A," consists of all pairs (t_1', t_2), where $t_1' \in A_1 \cap \hat{B}_1(A_2)$ and $t_2 \in T_2$. Note that this allows for the possibility that player 1 is wrong about player 2, but not that he is wrong about his own type.

The "local" belief operator $\hat{B}_i$ may be required to have the following logical properties. For all $A_j, F_j \subseteq T_j$:

B1 : $\hat{B}_i(T_j) = T_i$.
B2 : $\hat{B}_i(\emptyset) = \emptyset$.
B3 : $\hat{B}_i(A_j) \cap \hat{B}_i(F_j) \subseteq \hat{B}_i(A_j \cap F_j)$.
B4 : $A_j \subseteq F_j \Rightarrow \hat{B}_i(A_j) \subseteq \hat{B}_i(F_j)$.

The logical properties of belief can be related to restrictions on preferences. Morris (1997) shows that if preferences satisfy [P1]-[P4], Savage belief satisfies [B1]-[B4] and that if preferences satisfy [P1]-[P4] and

[P5*], strong belief satisfies [B1]-[B4]. In addition, if preferences satisfy [P1]-[P4], [P5*] and [P7], then Savage and strong belief are equivalent. If preferences satisfy [P6] and [P7], then Savage, strong*, strong and weak belief are equivalent.

Morris (1997, Lemma 1) shows that adding further syntactic restrictions yields the richer semantic structure for belief typically assumed by economists, which requires that possibility correspondences partition the state space.

The "composite" belief operator B_i inherits logical properties from the "local" belief operator $\hat{B}_i$. For instance, if $\hat{B}_i$ satisfies [B1], so that $\hat{B}_i(T_j) = T_i$, then it follows that $B_i(\Omega) = (T_i \cap \hat{B}_i(T_j)) \times T_j = T_i \times T_j \equiv \Omega$. Notice that if $\hat{B}_i$ satisfies [B1], $\hat{B}_i(T_j) = \hat{B}_i(\hat{B}_i(T_j))$.

Definition 6 *Consider an event $A \subseteq T_1 \times T_2$, where $A = A_1 \times A_2$. Denote the event "everyone believes A" by $B_*(A)$. Then*

$$B_*(A) = B_1(A) \cap B_2(A).$$

The event "everyone believes that everyone believes A" can then be denoted

$$[B_*]^2(A) \equiv B_*(B_*(A)) = B_1(B_*(A)) \cap B_2(B_*(A)) = B_1\left(B_1(A) \cap B_2(A)\right) \cap B_2\left(B_1(A) \cap B_2(A)\right),$$

and so on. The event "A is common belief" is thus

$$C(A) = \bigcap_{k=1}^{\infty} [B_*]^k(A).$$

Belief operators defined on players' type spaces thus provide us with a framework for expressing players' beliefs independent of any particular assumptions on preferences. Common belief is then defined simply as the iterated application of the appropriate belief operator.

4 Beliefs in Games

In this section, I investigate the epistemic foundations of iterative dominance. For now, I restrict attention to pure strategies, so that each player only plays pure strategies, and views the other player as doing so as well. In this framework, the event "player i is rational", as perceived by player j, corresponds to

$$R_i = \{t_i \in T_i : \ [g_i(f_i(t_i), f_j(t_j))]_{t_j \in T_j} \succeq_{t_i} [g_i(s_i, f_j(t_j))]_{t_j \in T_j} \ \forall \ s_i \in S_i\}.$$

That is, a rational type of player i will only take an action that is optimal in the sense of being at least as good as any other action available to him. A strategy s_i is then a best response for player i if there exists some type t_i of player i such that $f_i(t_i) = s_i$ and t_i is rational in the sense defined above. With this framework in hand, we can investigate strict and weak dominance.

Definition 7 *A strategy s_i is* dominated *if there exists $s_i' \in S_i$ such that*

$$g_i(s_i', s_j) > g_i(s_i, s_j) \; \forall \, s_j \in S_j.$$

Let U_i^∞ be the set of strategies of player i that survive iterated deletion of strictly dominated strategies, where dominance is by pure strategies. The first result is that with strong assumptions on what players know, rational players will play iteratively undominated strategies, regardless of the particular representation of their preferences.

Proposition 1 *In the 2-player, finite action game $< S_1, S_2, g_1, g_2 >$ let $(t_1, t_2) \in C(R_1 \times R_2)$, where belief is defined as Savage or strong* belief. Suppose that preferences satisfy [P1]-[P4] and [P5*]. Then, for $i = 1, 2$, $f_i(t_i) \in U_i^\infty$. When preferences satisfy [P7] in addition, the result also holds if $\hat{B}_1$ and $\hat{B}_2$ are defined as strong belief.*

When the notion of belief is weakened further, it is not always possible to rule out play of iteratively strictly dominated strategies. If players have MEU preferences, then common belief of rationality may still imply that they play iteratively dominated strategies.

Ryan (2000) argues that one reason to object to weak belief is that it sometimes fails to satisfy logical coherence, i.e., that it can fail [B2]. This can happen for MEU preferences if the intersection of the supports of the decision maker's beliefs is empty, so that no state is considered possible. The following example shows that, even when weak belief satisfies [B2], it can be too weak to rule out very much.

Example 1 Consider the following 2-player game.

		2: l	r
1	U	2,1	1,0
	D	1,2	2,1

Let $T_1 = \{\hat{t}_1, t_1'\}$ and $T_2 = \{\hat{t}_2, t_2', t_2''\}$. Suppose that players have MEU preferences, with utilities denoted by $\{\mathbf{u}_{t_i}\}_{t_i \in T_i}$, and belief sets $\{\Delta_{t_i}\}_{t_i \in T_i}$.

Now if player 2 is rational, then she will never play r, which is strictly dominated for her. Once r is eliminated, D is strictly dominated for player 1, so that the unique outcome that survives iterated deletion of strictly dominated strategies is (U, l). Let $R_2 = \{\hat{t}_2\}$ be the event "player 2 is rational"; then $f_2(\hat{t}_2) = l$. Suppose that $f_2(t_2') = f_2(t_2'') = r$ and $\Delta_{t_2} = \{\hat{p}_2\}$, where $\hat{p}_2(\hat{t}_1) = 1$ for all $t_2 \in T_2$. I am interested in when player 1 will prefer the act induced by D, $(1_{R_2}, 2_{-R_2})$, over the one induced by U, $(2_{R_2}, 1_{-R_2})$. Let

$f_1(\hat{t}_1) = D$, and $\Delta_{\hat{t}_1}$ be the convex hull of $\{\hat{p}_1, p_1'\}$, where the weights placed on the different states are:

	$\hat{t}_2$	t_2'	t_2''
$\hat{p}_1$	$\frac{1}{3}$	0	$\frac{2}{3}$
p_1'	$\frac{1}{3}$	$\frac{2}{3}$	0

Then at $(\hat{t}_1, \hat{t}_2)$, the only state that player 1 considers possible is $\hat{t}_2$, and the only state that player 2 considers possible is $\hat{t}_1$. Notice that although player 1 believes that player 2 is rational, the *maximum* weight he ever places on the event R_2 is $\frac{1}{3}$, whereas the *minimum* weight he places on its complement, which is *not* believed, is $\frac{2}{3}$. Hence

$$\mathbf{u}_{\hat{t}_1}(U) = \min_{p_1 \in \Delta_{\hat{t}_1}} [p_1(R_2)\, 2 + p_1(-R_2)\, 1] = \min_{p_1 \in \Delta_{\hat{t}_1}} [2\, p_1(R_2) + (1 - p_1(R_2))]$$
$$= 1 + \min_{p_1 \in \Delta_{\hat{t}_1}} p_1(R_2) = 1 + \frac{1}{3},$$

and

$$\mathbf{u}_{\hat{t}_1}(D) = \min_{p_1 \in \Delta_{\hat{t}_1}} [p_1(R_2)\, 1 + p_1(-R_2)\, 2] = 1 + \min_{p_1 \in \Delta_{\hat{t}_1}} p_1(-R_2) = 1 + \frac{2}{3},$$

so that $\mathbf{u}_{\hat{t}_1}(D) > \mathbf{u}_{\hat{t}_1}(U)$, and player 1 will (rationally) prefer playing strategy D to strategy U, despite weakly believing that player 2 is rational and will therefore not play r. Therefore, common weak belief of rationality does not rule out play of iteratively strictly dominated strategies. $\diamond$

When preferences are taken to satisfy admissibility, it is possible to rule out play of weakly dominated strategies. Consider two strategies $s_i, s_i' \in S_i$, such that $g_i(s_i', s_j) \geq g_i(s_i, s_j)\ \forall s_j \in S_j$, with strict inequality for some $s_j' \in S_j$. If there exists $t_j' \in T_j$ such that $f_j(t_j') = s_j'$, so that $g_i(s_i', f_j(t_j)) \geq g_i(s_i, f_j(t_j))\ \forall t_j \in T_j$, with strict inequality for t_j', admissibility implies $G_i(s_i') \succ_{t_i} G_i(s_i)$, so that type t_i of player i will not play the weakly dominated strategy, s_i. Assume from now on that every action is played by some type of each player.

When preferences are admissible, Savage belief is uninformative in that nothing non-trivial is ever believed (Morris (1997, Lemma 3)). We may therefore ask what the implications of common strong belief of rationality are for the play of undominated strategies.

Let S_i^∞ be the set of strategies which survive one round of deletion of weakly dominated strategies, followed by subsequent rounds of deletion of strictly dominated strategies, where domination is by pure strategies.

Proposition 2 *In the 2-player, finite action game $< S_1, S_2, g_1, g_2 >$ when preferences satisfy [P1] - [P5], if $(t_1, t_2) \in C(R_1 \times R_2)$, where belief is defined as strong belief, then, for $i = 1, 2$, $f_i(t_i) \in S_i^\infty$.*

Brandenburger (1992) considers the case of preferences which satisfy LEU. He defines the set of "permissible" strategies as those which are chosen if there is common first-order knowledge of rationality (common strong belief), and shows that this is equivalent to the set of strategies which survive the iterated deletion procedure above, where dominance is defined to include dominance by mixed strategies.

5 Equilibrium

This section considers equilibrium concepts for players in 2-person games.

Definition 8 *A strategy profile* $s^* = (s_1^*, s_2^*)$ *is a* pure-strategy Nash Equilibrium *if, for* $i = 1, 2$,

$$g_i(s_i^*, s_{-i}^*) \geq g_i(s_i', s_{-i}^*) \;\; \forall s_i' \in S_i.$$

I now show that rationality and mutual Savage or strong* belief of actions imply that players' choices will be a Nash equilibrium.

Proposition 3 *Let* $s^* = (s_1^*, s_2^*)$ *be a pair of actions. For* $i = 1, 2$, *let*

$$Q_i = \{t_i \in T_i : f_i(t_i) = s_i^*\}.$$

Suppose for some $t^* = (t_1^*, t_2^*)$ *that* $t^* \in R_1 \times R_2$, *and* $t^* \in B_*(Q_1 \times Q_2)$, *where* $\hat{B}_1$ *and* $\hat{B}_2$ *denote Savage or strong* belief. Suppose that preferences satisfy [P1] - [P4] and [P5*]. Then* s^* *is a Nash equilibrium. If preferences satisfy [P1]-[P4] and [P5], then strong belief is sufficient to ensure that* s^* *is a Nash equilibrium.*

Strong* and Savage belief thus serve as benchmark notions. In proposition 3, belief of actions is sufficiently strong to rule out either player deviating unilaterally given what the other player is playing. However, once we relax belief of actions to mean weak belief, we can no longer rule out non-Nash behavior.

Example 2 Consider the following 2-player game.

		2: l	r
1	t	1, 3	2, 4
	b	2, 2	1, 1

Players are assumed to have MEU preferences $\{\mathbf{u}_{t_i}\}_{t_i \in T_i}$, $\{\Delta_{t_i}\}_{t_i \in T_i}$. Let $[b]$ denote the event that player 1 plays the strategy b, and let $[r]$ denote the event that player 2 plays the strategy r. Strong or Savage belief of $[r]$ implies that, for any $t_1 \in \hat{B}_1([r]), \delta_1([r]) = 1 \; \forall \; \delta_1 \in \Delta_{t_1}$. Clearly, this implies that player 1 would prefer the act induced by t, $(1_{-[r]}, 2_{[r]})$ over the one induced by b, which is $(2_{-[r]}, 1_{[r]})$, so that it is not possible to support play of b when player 1 strongly believes that player 2 is playing r. However, suppose that type t_1' of player 1 weakly believes that player 2 is playing r. Weak belief imposes only the restriction that, for all $t_2 \notin [r]$, there exists $p_1 \in \Delta_{t_1'}$ such that $p_1(t_2) = 0$. Suppose $\min_{p_1 \in \Delta_{t_1'}} p_1([r]) = 1/4$ and $\min_{p_1 \in \Delta_{t_1'}} p_1(-[r]) = 1/3$. Then

$$\mathbf{u}_{t_1}(t) = \min_{p_1 \in \Delta_{t_1}} [1 + p_1([r])] = 1 + \frac{1}{4} < \mathbf{u}_{t_1'}(b) = \min_{p_1 \in \Delta_{t_1'}} [1 + p_1(-[r])] = 1 + \frac{1}{3}.$$

Thus type t_1' of player 1 prefers playing strategy b, even though he weakly believes player 2 is playing r. Similarly, for type t_2' of player 2, suppose that $\min_{p_2 \in \Delta_{t_2'}} p_2(-[b]) = 9/16$. Then

$$\mathbf{u}_{t_2'}(l) = \frac{41}{16} < u_{t_2'}(r) = \frac{43}{16} .$$

So player 2 prefers playing strategy r despite weakly believing that player 1 plays b. Hence weak belief of actions is not sufficient to rule out non-Nash behavior. ◇

The logical next step is to consider equilibrium in mixed strategies. There are two accepted interpretations of mixed strategy Nash Equilibrium. The traditional interpretation is that players actually mix according to the equilibrium strategies; the second interpretation holds that player 1's mixed strategy represents not his actual action but player 2's beliefs about what pure strategy player 1 is going to pick. As is well known, however, this equivalence between beliefs and strategies breaks down when players do not have expected utility preferences: when beliefs are not represented by a single probability measure, then a player's mixed strategy has no ready interpretation as the other player's belief over her possible pure strategies.

In the present context, the appropriate counterpart to an equilibrium in beliefs may be defined following Lo (1996) in terms of preferences.

Definition 9 $\{\succeq_{t_i}, \succeq_{t_j}\}$ *is a* Nash Equilibrium in preferences *if*

1. *there exists $\Gamma_i \subseteq S_i$ such that the event $-F_i$, where $F_i = f_i^{-1}(\Gamma_i)$, is null with respect to $\succeq_{t_j}$, $j \neq i$; and*

2. *for all $s_i \in \Gamma_i$, $G_i(s_i) \succeq_{t_i} G_i(s_i')$ $\forall$ $s_i' \in S_i$.*

In words, the preferences form a Nash Equilibrium if the event that each player is irrational is null from the point of view of the other player, and each player's action is optimal given her preferences.

Proposition 4 *Suppose for some $t^* = (t_1^*, t_2^*)$ that players are rational and that this is mutual belief, so that $t^* \in R_1 \times R_2$, and $t^* \in B_*(R_1 \times R_2)$. Let player 1's belief regarding player 2 be $F_2 \subseteq T_2$, so that $t_1^* \in \hat{B}_1(F_2)$; similarly, let $t_2^* \in \hat{B}_2(F_1)$. Suppose that $t^* \in B_*(\hat{B}_1(F_2) \times \hat{B}_2(F_1))$, i.e., that the beliefs are mutually believed. If the belief operator is normal, then $\{\succeq_{t_1^*}, \succeq_{t_2^*}\}$ is a Nash Equilibrium in preferences.*

This result can be viewed as a general analogue to Theorem A in Aumann and Brandenburger (1995). It is chiefly remarkable for what it does not say: because it imposes no explicit structure on preferences, and very little structure on beliefs, it does not permit us to evaluate the different notions of equilibrium that rely on specific representations of preferences. Moreover, it does not solve the problem of mixed versus pure strategies: if all definitions were in terms of mixtures, then the analogue of the proposition with preferences and belief defined over mixed strategies would go through[2] .

6 Conclusion and Extensions

This paper has offered a unified language for thinking about interactive beliefs in the context of non-expected utility maximisers. By exploring a minimal preference structure compatible with games from a subjective viewpoint, the paper provides a framework within which extensions of single-person decision theory to games can be compared and evaluated. By considering the relation of beliefs to preferences, the paper makes it possible to identify the factors that drive notions of dominance and equilibrium. The paper also identifies some of the weaknesses of the existing extensions of non-expected utility theory to games.

The paper provides a foundation for future work relating preferences and logical properties of belief in games. Work remains to be done on relating the minimal dominance notions explored here to general notions of rationalizability and equilibrium, and on providing foundations. Also, work by Morris (1996) shows how substantive properties of belief may be related to preferences; the implications of these properties for interactive beliefs in games remain to be explored. The extension to n players is also a subject for future work.

References

[1] Aumann, R.J. (1976), "Agreeing to Disagree," *Annals of Statistics* 4, 1236-1239.

[2] Aumann, R.J. and A. Brandenburger (1995), "Epistemic Conditions for Nash Equilibrium," *Econometrica* 63, 1161-1180.

[3] Bernheim, D. (1984), "Rationalizable Strategic Behavior", *Econometrica* 52, 1007-1028.

[4] Blume, L., A. Brandenburger and E. Dekel (1991), "Lexicographic Probabilities and Choice Under Uncertainty," *Econometrica* 59, 61-80.

[5] Blume, L., A. Brandenburger and E. Dekel (1991), "Lexicographic Probabilities and Equilibrium Refinements," *Econometrica* 59, 81-98.

[6] Brandenburger, A. (1992), "Lexicographic Probabilities and Iterated Admissibility," in P.S. Dasgupta *et al.*, eds. *The Economic Analysis of Markets and Games.* Cambridge, MA: MIT Press.

[7] Dow, J. and S. Werlang (1994), "Nash Equilibrium under Knightian Uncertainty: Breaking Down Backward Induction," *Journal of Economic Theory* 64, 305-324.

[2] Proposition 6 in Lo (1996) is a special case that is very similar in spirit to this result, but employs mixed strategies.

[8] Eichberger, J. and D. Kelsey (1998), "Non-additive Beliefs and Strategic Equilibria," emphGames and Economic Behavior, 30, 183-215.

[9] Ellsberg, D. (1961), "Risk, Ambiguity, and the Savage axioms", *Quarterly Journal of Economics* 75, 643-669.

[10] Epstein, L. (1997), " Preference, Rationalizability and Equilibrium," *Journal of Economic Theory*, 73, 1-29.

[11] Gilboa, I. (1987), "Expected Utility Theory with Purely Subjective Non-additive Probabilities," *Journal of Mathematical Economics*, 16, 65-88.

[12] Gilboa, I. and D. Schmeidler (1989), "Maxmin Expected Utility with Non-unique Prior," *Journal of Mathematical Economics*, 18, 141-153.

[13] Klibanoff, P. (1996), "Uncertainty, Decision and Normal Form Games," mimeo, Northwestern University.

[14] Lo, K.C. (1996), "Equilibrium in Beliefs under Uncertainty," *Journal of Economic Theory* 71, 443-484.

[15] Marinacci, M. (1996), "Ambiguous Games," forthcoming in *Games and Economic Behavior.*

[16] Morris, S. (1996), "The Logic of Belief and Belief Change: A Decision Theoretic Approach," *Journal of Economic Theory* 69, 1-23.

[17] Morris, S. (1997), "Alternative Definitions of Knowledge," in M.O.L. Bacharach *et al.* eds., *Epistemic Logic and the Theory of Games and Decisions.* Dordrecht, The Netherlands: Kluwer Academic.

[18] Mukerji, S. (1997), "Understanding the Nonadditive Probability Decision Model," *Economic Theory* 9, 23-46.

[19] Mukerji, S. and H.S. Shin (1997), "Equilibrium Departures from Common Knowledge in Games with Non-Additive Expected Utility," mimeo, University of Southampton.

[20] Pearce, D. (1984), "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica* 52, 1029-1050.

[21] Ryan, M.J. (1998), "Rational Belief Revision by Uncertainty-Averse Decision Makers," mimeo, University of Auckland.

[22] Ryan, M.J. (2000), "What do Uncertainty-Averse Decision Makers Believe?," mimeo, Australian National University.

[23] Savage, L. (1954), *The Foundations of Statistics*, John Wiley and Sons.

[24] Schmeidler, D. (1989), "Subjective Probability and Expected Utility Without Additivity," *Econometrica* 57, 571-587.