

Nash Equilibrium without Mutual Knowledge of Rationality¹

Kin Chung Lo

Department of Economics, University of Toronto,
Toronto, Ontario, Canada M5S 1A1

July, 1995

Abstract

This paper defines an equilibrium concept for general preferences in two person normal form games. It collapses to Nash Equilibrium when preferences are represented by the expected utility model. An important characteristic of the equilibrium concept is that player i does not necessarily know that player j is rational, but she views rationality as infinitely more likely than irrationality. For suitable models of preferences, the equilibrium concept predicts that a player will take a “cautious” strategy that is not a best response in any Nash Equilibrium. *Journal of Economic Literature* Classification Numbers: C72, D81.

¹I especially thank Professor Larry G. Epstein for pointing out this topic, and for providing supervision and encouragement. Remaining errors are my responsibility.

1. INTRODUCTION

The expected utility model axiomatized by Savage (1954) has been the most popular model for studying decision making under uncertainty. It is also almost universally used in game theory. Under the assumption that players' preferences are represented by expected utility functions, a large number of equilibrium concepts have been developed, the central one being Nash Equilibrium. Adopting the view that choosing a strategy in a game is a decision problem under uncertainty, Aumann and Brandenberger (1991) interpret a Nash Equilibrium as a collection of beliefs held by the players about opponents' strategy choices. They demonstrate formally that in a two person normal form game, if the structure of the game and players' beliefs and rationality are mutual knowledge, then the beliefs profile constitutes a Nash Equilibrium.²

Game theorists have been questioning the assumption of knowledge of rationality:

“Common knowledge of rationality seems a very strong assumption. Even mutual knowledge of rationality is quite strong. A player may be rational himself, but how can he know for sure that another person is?” (Aumann (1992, p.219)).

Irrationality in game theory is an important topic for research. To relax the assumption of mutual knowledge of rationality, one approach that has been taken is to assume that irrationality is very unlikely but it is not a null event. When players are expected utility maximizers, this means that a player attaches probability “very close” (but not equal) to 1 to the event that the other player is rational. It is well recognized that this can make a very big difference in the outcome of a game.

In this paper, I adopt the above idea, but consider more general preferences. Let me make use of Example 1 to provide the motivation. For all games presented in this paper, player i (male) is the row player and player j (female) is the column player. Payoffs are in terms of utility.

Example 1

	L	R
l	10 million, 10	-10 million, 9
m	9 million, 10	8 million, 9
n	8 million, 10	9 million, 9
o	-10 million, 10	400 million, 9

²Aumann and Brandenberger (1991) also point out that it is natural for the conditions to be sufficient but not necessary because nothing can prevent the beliefs of the players from forming a Nash Equilibrium “accidentally”.

If player i knows that player j is rational, then i knows that j will play L and therefore i should play l . In fact, the strategy profile $\{l, L\}$ constitutes the unique Nash Equilibrium.

However, it is intuitive that player i may play m . The informal justification is as follows: player i “expects” that player j is rational and therefore she will play L . If player i plays o , he will receive -10 million. Therefore, even though o gives a very high payoff in case player j plays R , this does not induce player i to play o . We only need player i to believe that it is more likely that player j is rational to justify that m is strictly better than n . Player i also strictly prefers to play m rather than l because although player i expects that player j is rational and therefore she will play L , he does not “know for sure” that this is really the case. If player i plays l but player j plays R , player i will receive -10 million. To avoid completely the possibility of receiving the misfortune of -10 million, player i therefore decides to play m . By doing so, he knows that he can receive at least 8 million regardless of what player j is going to do.

This example also conveys the following important message: note that player i will not play m even if we do not employ Nash Equilibrium as the solution concept but retain the assumption that player i is an expected utility maximizer. If player i is an expected utility maximizer, l (o) is strictly better than m if player i attaches probability at least (at most) 0.95 to the event that player j will play L . Therefore, his best response is either l or o , but never m . This demonstrates the necessity to define a version of Nash Equilibrium that allows for more general preferences. ■

Motivated by Example 1, I do not start by restricting myself to a specific class of preferences. An equilibrium concept which is expressed only in terms of given preferences for players is provided first. Therefore a modeller who wants to make use of the equilibrium concept has complete freedom to adopt any class of preferences (provided that it ensures the existence of an equilibrium). Following the traditional approach of relaxing the assumption of mutual knowledge of rationality, my equilibrium concept allows irrationality to be a non-null event. Its distinctive feature is that given any two strategies of player j which i thinks that j may use, i believes that it is infinitely more likely that j will pick the strategy which she strictly prefers. Then I proceed to look for models of preferences that can be adopted to the equilibrium concept to deliver the desired prediction in Example 1.

Recently, a number of papers have formulated generalizations (or refinements) of Nash Equilibrium in normal form games by allowing players’ preferences to deviate from the expected utility model. They include Blume *et al.* (1991b), Crawford (1990), Dekel *et al.* (1991), Dow and Werlang (1994), Klibanoff (1993), Lo (1995) and Mukerji (1994). It is therefore important to differentiate

this paper from the above. The first difference is motivation. The motivation of this paper directly comes from modelling beliefs in irrationality in game theory. This is only shared by Blume *et al.* (1991b) and Mukerji (1994). In fact, the spirit of the equilibrium concept in this paper is similar to those in Blume *et al.* (1991b). I will explain in details the difference between my equilibrium concept and theirs. The motivation of all the other papers comes from the Allais and Ellsberg Paradoxes in single person decision theory. As a result, considerations of uncertainty about opponents' rationality is absent (or at least not the main focus) in their equilibrium concepts. The second difference is prediction, which can be highlighted using Example 1. My equilibrium concept satisfies the following properties:

1. There exist models of preferences that can be adopted to the equilibrium concept to predict that player i will play m .
2. There does not exist a model of preferences that can be adopted to the equilibrium concept to predict that player i will play n or o .

Nash Equilibrium and the equilibrium concepts in Blume *et al.* (1991b), Crawford (1990), Dekel *et al.* (1991) and Lo (1995) violate 1. On the other hand, those in Dow and Werlang (1994), Klibanoff (1993) and Mukerji (1994) violate 2. Moreover, my equilibrium concept can be regarded as a refinement of those in Dow and Werlang (1994) and Mukerji (1994), which is another contribution of this paper.

The paper is organised as follows. Since I also adopt the view that choosing a strategy in a game is a decision problem under uncertainty, I first consider single person decision theory in section 2. The theory is adapted to the context of normal form games in Section 3, where the equilibrium concept is presented. Section 4 demonstrates formally that my equilibrium concept, when specialized to the multiple priors model of Gilboa and Schmeidler (1989), is a refinement of those in Dow and Werlang (1994) and Mukerji (1994). Section 5 concludes.

2. SINGLE PERSON DECISION THEORY

Before I present the equilibrium concept, some preliminaries on single person decision theory are required. For any topological space Y , adopt the Borel σ -algebra Σ_Y and denote by $M(Y)$ the set of all Borel probability measures over Y with finite supports. Let (X, Σ_X) be the space of outcomes and (Ω, Σ_Ω) the space of uncertainty. For the purpose of this paper, I assume that $X = \mathbf{R}$ and Ω is a finite set. Let \mathcal{F} be the set of all functions from Ω to $M(X)$. That is, \mathcal{F} is the

set of two-stage, horse-race/roulette-wheel acts, as in Anscombe and Aumann (1963). f is called a constant act if $f(\omega) = p \quad \forall \omega \in \Omega$; such an act involves (probabilistic) risk but no uncertainty. For notational simplicity, I also use $p \in M(X)$ to denote the constant act that yields p in every state of the world, $x \in X$ the degenerate probability measure on x and $\omega \in \Omega$ the event $\{\omega\} \in \Sigma_\Omega$. The primitive \succeq is a weak preference ordering over acts. The relations of strict preference and indifference are denoted by \succ and \sim respectively.

There are several choice theoretic notions regarding \succeq that will be relevant. Say that \succeq is *monotonic* if for all $f, g \in \mathcal{F}$,

$$f(\omega) \succeq g(\omega) \quad \forall \omega \in \Omega \Rightarrow f \succeq g \quad \text{and} \quad f(\omega) \succ g(\omega) \quad \forall \omega \in \Omega \Rightarrow f \succ g.$$

Throughout this paper, assume that \succeq is monotonic and for all $x^*, x \in X$, $x^* \succ x$ if and only if $x^* > x$.

The next two notions are from Savage (1954). Say that an event T is \succeq -*null* if for all $f, f', g \in \mathcal{F}$,

$$\begin{bmatrix} f(\omega) & \text{if } \omega \in T \\ g(\omega) & \text{if } \omega \notin T \end{bmatrix} \sim \begin{bmatrix} f'(\omega) & \text{if } \omega \in T \\ g(\omega) & \text{if } \omega \notin T \end{bmatrix}.$$

In words, an event T is \succeq -null if the decision maker does not care about payoffs in states belonging to T . This can be interpreted as the decision maker knows (or believes) that T can never happen.

Given any two events A and B , say that A is \succeq -*more likely than* B if for all $x^*, x \in X$ with $x^* \succ x$,

$$\begin{bmatrix} x^* & \text{if } \omega \in A \\ x & \text{if } \omega \notin A \end{bmatrix} \succeq \begin{bmatrix} x^* & \text{if } \omega \in B \\ x & \text{if } \omega \notin B \end{bmatrix}.$$

In words, A is \succeq -more likely than B if the decision maker prefers to bet on A rather than on B .

We also need the following less standard notion:

Definition 1. A is \succeq -*infinitely more likely than* B if for all $x^*, x, y \in X$ with $x^* \succ x$,

$$\begin{bmatrix} x^* & \text{if } \omega \in A \\ x & \text{if } \omega \notin A \end{bmatrix} \succ \begin{bmatrix} y & \text{if } \omega \in B \\ x & \text{if } \omega \notin B \end{bmatrix}. \quad (1)$$

The interpretation of the \succeq -infinitely more likely than relation is as follows: A is \succeq -infinitely more likely than B if the decision maker strictly prefers to bet on A rather than on B and moreover, increasing the payoff by any extent for the act $\begin{bmatrix} y & \text{if } \omega \in B \\ x & \text{if } \omega \notin B \end{bmatrix}$ in all states belonging to B does not induce his strict preference to change. The \succeq -infinitely more likely than relation possesses some natural properties. It is irreflexive, asymmetric and transitive. Moreover, \succeq is monotonic implies

that if A is \succeq -infinitely more likely than B , then A is \succeq -infinitely more likely than any subset of B .

I should point out that the idea of introducing an \succeq -infinitely more likely than relation is not new. Blume *et al.* (1991a, Definition 5.1) also define an \succeq -infinitely more likely than relation in the context of single person decision theory. According to their definition, for any two disjoint events A and B , A is \succeq -infinitely more likely than B if for all $f, g, h \in \mathcal{F}$,

$$\Rightarrow \left[\begin{array}{l} f(\omega) \text{ if } \omega \in A \\ h(\omega) \text{ if } \omega \notin A \end{array} \right] \succ \left[\begin{array}{l} g(\omega) \text{ if } \omega \in A \\ h(\omega) \text{ if } \omega \notin A \end{array} \right] \\ \Rightarrow \left[\begin{array}{l} f(\omega) \text{ if } \omega \in A \\ \check{h}(\omega) \text{ if } \omega \in B \\ h(\omega) \text{ if } \omega \notin A \cup B \end{array} \right] \succ \left[\begin{array}{l} g(\omega) \text{ if } \omega \in A \\ \check{h}(\omega) \text{ if } \omega \in B \\ h(\omega) \text{ if } \omega \notin A \cup B \end{array} \right] \quad \text{for all } \check{h}, \check{h} \in \mathcal{F}. \quad (2)$$

The main difference between (1) and (2) is that to test whether A is \succeq -infinitely more likely than B , (1) only compares binary (two outcome) acts, but (2) considers more general acts.

Given the \succeq -infinitely more likely than relation defined by Blume *et al.* (1991a), it is important to explain the value added of Definition 1. Note that when $A \cup B = \Omega$, A is \succeq -infinitely more likely than B according to (2) only if for all $x^*, x, y \in X$ with $x^* \succ x$,

$$\left[\begin{array}{l} x^* \text{ if } \omega \in A \\ x \text{ if } \omega \in B \end{array} \right] \succ \left[\begin{array}{l} x \text{ if } \omega \in B \\ x \text{ if } \omega \in A \end{array} \right] \Rightarrow \left[\begin{array}{l} x^* \text{ if } \omega \in A \\ y \text{ if } \omega \in B \end{array} \right] \succ \left[\begin{array}{l} x^* \text{ if } \omega \in B \\ x \text{ if } \omega \in A \end{array} \right]. \quad (3)$$

That is, the decision maker strictly prefers to bet on A rather than on B and moreover, decreasing the payoff by any extent for the act $\left[\begin{array}{l} x^* \text{ if } \omega \in A \\ y \text{ if } \omega \in B \end{array} \right]$ does not induce his strict preference to change. There is intuition for \succeq to obey (1) but violate (3). When a decision maker expects that A will happen, giving him an act that delivers a big prize if A does not happen does not make him excited. On the other hand, giving him an act that delivers a big misfortune in case A does not happen may make him feel uncomfortable. That is, violation of (3) may be justified by the psychological desire of the decision maker to rule out completely the possibility of a misfortune. In fact, if we require player i to believe that it is infinitely more likely that player j will pick the strategy that she strictly prefers in Example 1, then it is necessary for the preference ordering \succeq of player i to violate (3) in order to predict that he will play m . To see this, let $\Omega = \{L, R\}$ be the state space for player i , $A = L$ and $B = R$. Player i believes that it is \succeq -infinitely more likely that player j will pick the strategy which she strictly prefers if and only if L is \succeq -infinitely more likely than R . If (3) were satisfied, we would have $l \succ m$. Therefore, using the \succeq -infinitely more likely than relation in Blume *et al.* (1991a) cannot predict that player i plays m .

The next task is to identify models of preferences where introducing Definition 1 will make a difference. When preferences are represented by the expected utility model, A is \succeq -infinitely

more likely than B if and only if A is not $\underline{\succ}$ -null and B is $\underline{\succ}$ -null. Therefore given the notion of $\underline{\succ}$ -null event, there is no need to introduce the $\underline{\succ}$ -infinitely more likely than relation. In fact, this continues to hold for all preferences that are strictly monotonic in all non- $\underline{\succ}$ -null events and satisfy an “intermediate value property”. Say that $\underline{\succ}$ is *strictly monotonic in all non- $\underline{\succ}$ -null events* if for all $x^*, x, y \in X$ with $x^* \succ x$ and for all non- $\underline{\succ}$ -null events T ,³

$$\begin{bmatrix} x^* & \text{if } \omega \in T \\ y & \text{if } \omega \notin T \end{bmatrix} \succ \begin{bmatrix} x & \text{if } \omega \in T \\ y & \text{if } \omega \notin T \end{bmatrix}. \quad (4)$$

Say that $\underline{\succ}$ satisfies the *intermediate value property* if for all non- $\underline{\succ}$ -null events T and for all $x^*, x \in X$ with $x^* \succ x$, there exists $y \succ x$ such that

$$\begin{bmatrix} x^* & \text{if } \omega \in T \\ x & \text{if } \omega \notin T \end{bmatrix} \underline{\succ} y. \quad (5)$$

Proposition 1. *Suppose that $\underline{\succ}$ is strictly monotonic in all non- $\underline{\succ}$ -null events and satisfies the intermediate value property. Then given any two events A and B , A is $\underline{\succ}$ -infinitely more likely than B if and only if A is not $\underline{\succ}$ -null and B is $\underline{\succ}$ -null.*

Proof:

\Rightarrow

Suppose that A is $\underline{\succ}$ -infinitely more likely than B . Then for all $x^*, x \in X$ with $x^* \succ x$,

$$\begin{bmatrix} x^* & \text{if } \omega \in A \\ x & \text{if } \omega \notin A \end{bmatrix} \succ \begin{bmatrix} x & \text{if } \omega \in B \\ x & \text{if } \omega \notin B \end{bmatrix} \sim \begin{bmatrix} x & \text{if } \omega \in A \\ x & \text{if } \omega \notin A \end{bmatrix}.$$

This implies that A is not $\underline{\succ}$ -null.

Next I show that B is $\underline{\succ}$ -null. Suppose that B were not $\underline{\succ}$ -null. Then for each $x^* \succ x$, there exists $y \succ x$ such that

$$\begin{bmatrix} x^* & \text{if } \omega \in B \\ x & \text{if } \omega \notin B \end{bmatrix} \underline{\succ} \begin{bmatrix} y & \text{if } \omega \in A \\ y & \text{if } \omega \notin A \end{bmatrix} \underline{\succ} \begin{bmatrix} y & \text{if } \omega \in A \\ x & \text{if } \omega \notin A \end{bmatrix}.$$

The first weak preference relation is due to the intermediate value property. The last weak preference relation is due to the assumption that $\underline{\succ}$ is monotonic. The above establishes a contradiction to the hypothesis that A is $\underline{\succ}$ -infinitely more likely than B .

\Leftarrow

³Savage’s axiom P3 is stronger than (4) in the sense that the constant act y is replaced by any (not necessarily constant) act.

Suppose A is not \succeq -null and B is \succeq -null, then for all $x^*, x, y \in X$ with $x^* \succ x$,

$$\begin{bmatrix} x^* & \text{if } \omega \in A \\ x & \text{if } \omega \notin A \end{bmatrix} \succ \begin{bmatrix} x & \text{if } \omega \in A \\ x & \text{if } \omega \notin A \end{bmatrix} \sim \begin{bmatrix} x & \text{if } \omega \in B \\ x & \text{if } \omega \notin B \end{bmatrix} \sim \begin{bmatrix} y & \text{if } \omega \in B \\ x & \text{if } \omega \notin B \end{bmatrix}.$$

The first strict preference relation is due to the fact that \succeq is strictly monotonic in all non- \succeq -null events. The last indifference relation is due to the fact that B is \succeq -null. This completes the proof that A is \succeq -infinitely more likely than B . ■

It is intuitive that \succeq may violate (4) or (5) if T is an event which is “infinitely unlikely” to happen. (That is, there exists an event A such that A is \succeq -infinitely more likely than T .) In fact, such models of preferences are readily available and have received considerable attention in the decision theory literature. A brief review of some of these models are now provided.

Gilboa and Schmeidler (1989) propose the multiple priors model: there exists an *affine* function $u : M(X) \rightarrow \mathbf{R}$ and a unique, nonempty, closed and convex set Δ of probability measures on Ω such that for all $f, g \in \mathcal{F}$,

$$f \succeq g \Leftrightarrow \min_{p \in \Delta} \int u \circ f dp \geq \min_{p \in \Delta} \int u \circ g dp. \quad (6)$$

It is convenient to interpret Δ as representing the beliefs of the decision maker over Ω . This model violates (4) but satisfies (5). An event T is \succeq -null if and only if $p(T) = 0 \quad \forall p \in \Delta$. \succeq is strictly monotonic in T if and only if $p(T) > 0 \quad \forall p \in \Delta$. Given any two events A and B , A is \succeq -infinitely more likely than B if and only if $p(A) > 0 \quad \forall p \in \Delta$ and $\exists p \in \Delta$ such that $p(B) = 0$. Therefore A is \succeq -infinitely more likely than B implies that \succeq is not strictly monotonic in B , but B is not necessarily a \succeq -null event. The characterization of \succeq -null event and the \succeq -infinitely more likely than relation in terms of Δ motivates the following definition.⁴

Definition 2. The *extended support* of Δ is $\{\omega \in \Omega \mid \exists p \in \Delta \text{ such that } p(\omega) > 0\}$. The *support* of Δ is $\{\omega \in \Omega \mid p(\omega) > 0 \quad \forall p \in \Delta\}$.

The extended support of Δ is nonempty. A state $\omega \in \Omega$ is not \succeq -null if and only if it is in the extended support of Δ . The support of Δ is may be empty. Every state in the support of Δ is \succeq -infinitely more likely than every state in the complement of the support of Δ .

Blume *et al.* (1991a) propose the lexicographic expected utility model: there exists an affine function $u : M(X) \rightarrow \mathbf{R}$ and a collection $\{p^k\}_{k=1}^K$ of probability measures on Ω such that for all

⁴The definition of extended support in this paper is the same as that in Dow and Werlang (1994). The definition of support in this paper is different from theirs.

$f, g \in \mathcal{F}$,

$$f \succeq g \Leftrightarrow \left[\int u \circ f dp^k \right]_{k=1}^K \geq_L \left[\int u \circ g dp^k \right]_{k=1}^K \quad (7)$$

where \geq_L is the lexicographic ordering in Euclidean space. Again, $\{p^k\}_{k=1}^K$ can be interpreted as representing the beliefs of the decision maker over Ω . Unlike the multiple priors model, this model satisfies (4) but violates (5). Unfortunately, the two \succeq -infinitely more likely than relations defined in (1) and (2) coincide when they are applied to this model. This is the reason that the equilibrium notion defined in the next section, when specialized to this model of preferences, cannot deliver the prediction that player i will play m in Example 1.

Finally, I provide an example of \succeq where (4) is satisfied, (5) is violated and the equivalence of (1) and (2) breaks down. Consider the following lexicographic multiple priors utility function: there exists an affine function $u : M(X) \rightarrow \mathbf{R}$ and a collection $\{\Delta^k\}_{k=1}^K$ of nonempty, closed and convex sets of probability measures on Ω such that for all $f, g \in \mathcal{F}$,

$$f \succeq g \Leftrightarrow \left[\min_{p^k \in \Delta^k} \int u \circ f dp^k \right]_{k=1}^K \geq_L \left[\min_{p^k \in \Delta^k} \int u \circ g dp^k \right]_{k=1}^K. \quad (8)$$

Furthermore, let $\{\Delta^k\}_{k=1}^K$ have the property that for each $\omega \in \Omega$, there exists $1 \leq k \leq K$ such that $\omega \in$ extended support of Δ^k if and only if there exists $1 \leq k' \leq K$ such that $\omega \in$ support of $\Delta^{k'}$. This property ensures that \succeq satisfies (4). The fact that (1) and (2) are not equivalent and \succeq violates (5) are easily verified.

3. EQUILIBRIUM CONCEPT

Let me first describe how choosing a strategy in a game can be viewed as a decision problem under uncertainty. Player i 's finite strategy space is S_i with typical element s_i . The game specifies an outcome function $g_i : S_i \times S_j \rightarrow X$ for player i which is assumed to be common knowledge. Since player i is uncertain about the strategy choice of player j , the relevant state space for player i is S_j . Consistent with the above single person decision theory, player i has a preference ordering \succeq_i over the set of acts defined on S_j . Each strategy s_i of player i can be identified as an act over S_j as follows: if player i chooses s_i and the true state is s_j , player i receives the outcome $g_i(s_i, s_j)$. His objective is to choose $s_i \in S_i$ such that $s_i \succeq_i \hat{s}_i \quad \forall \hat{s}_i \in S_i$.

Now I am in a position to state my equilibrium concept that can possibly deliver the desired prediction in Example 1.

Definition 3. $\{\succsim_i, \succsim_j\}$ is a *Cautious Nash Equilibrium* if $s_i \succsim_i \hat{s}_i$ for all $s_i \in S_i$ that are not \succsim_j -null and for all $\hat{s}_i \in S_i$ not \succsim_j -infinitely more likely than s_i ; and similarly for j .

When \succeq_i and \succeq_j are expected utility preferences, Definition 3 reduces to the usual definition of Nash Equilibrium.

Cautious Nash Equilibrium can never predict that player i will play n or o in Example 1, regardless of the model of preferences adopted. The reason is that L strictly dominates R . Therefore we must have $L \succ_j R$ for any \succ_j . This implies that if $\{\succeq_i, \succeq_j\}$ is a Cautious Nash Equilibrium, L must be \succeq_i -infinitely more likely than R . This in turn implies that $m \succ_i n$ and $l \succ_i o$.

It is possible that in a Cautious Nash Equilibrium, player i will play m in Example 1. The reason is that the equilibrium concept only requires the strategy L to be \succeq_i -infinitely more likely than the strategy R ; it does not require R to be \succeq_i -null. That is, it does not require player i to rule out completely the possibility that player j is irrational and therefore that she will play R . Proposition 1 also makes clear that Cautious Nash Equilibrium predicts the prudent behavior of player i only if \succeq_i violates (4) and/or (5). Otherwise, L is \succeq_i -infinitely more likely than R implies that R is \succeq_i -null. If player i rules out completely the possibility that player j will play R , then his unique best response is l .

To provide a concrete illustration of Cautious Nash Equilibrium and demonstrate that there exist models of preferences that can be adopted to the equilibrium concept to predict that player i will play m in Example 1, let me restate Definition 3 in terms of the multiple priors model. First adapt the multiple priors model from the context of single person decision making to the context of normal form games. Player i 's induced preference ordering \succeq_i over S_i is represented by the following utility function: there exists an *affine* function $\hat{u}_i : M(X) \rightarrow \mathbf{R}$ and a nonempty, closed and convex set of probability measures B_i over S_j such that for all $s_i, \hat{s}_i \in S_i$,

$$s_i \succeq_i \hat{s}_i \Leftrightarrow \min_{b_i \in B_i} u_i(s_i, b_i) \geq \min_{b_i \in B_i} u_i(\hat{s}_i, b_i)$$

where $u_i(s_i, b_i) \equiv \sum_{s_j \in S_j} \hat{u}_i(g_i(s_i, s_j))b_i(s_j)$.

Cautious Nash Equilibrium, when restated in terms of the multiple priors model, is the following:

Definition 4. $\{B_i, B_j\}$ is a *Cautious Nash Equilibrium* if the following conditions are satisfied:⁵

1. Every $s_i \in$ support of B_j satisfies

$$\min_{b_i \in B_i} u_i(s_i, b_i) \geq \min_{b_i \in B_i} u_i(\hat{s}_i, b_i) \quad \forall \hat{s}_i \in S_i.$$

⁵Strictly speaking, I should state the equilibrium as $\{(B_i, u_i), (B_j, u_j)\}$ in order to be consistent with the notation $\{\succeq_i, \succeq_j\}$ in Definition 3. Here, I conform to the usual way of defining Nash Equilibrium, which is in terms of a pair of probability measures.

2. Every $s_i \in$ extended support of B_j satisfies

$$\min_{b_i \in B_i} u_i(s_i, b_i) \geq \min_{b_i \in B_i} u_i(\hat{s}_i, b_i) \quad \forall \hat{s}_i \notin \text{support of } B_j;$$

and similarly for j .

The following constitutes a Cautious Nash Equilibrium for the game in Example 1:

$$B_i = \{b_i \in M(\{L, R\}) \mid 0.9 \leq b_i(L) \leq 1\} \quad \text{and} \quad B_j = \{m\}.$$

Given the beliefs B_i of player i , L is \succeq_i -infinitely more likely than R . However, R is not a \succeq_i -null event. This is the reason that player i wants to be “cautious”. Given the beliefs B_i of player i , the unique best response is m . Finally, note that although R is not a \succeq_i -null event, \succeq_i is not strictly monotonic in R . This confirms the discussion in section 2 that in order to predict that player i will play m in equilibrium, if \succeq_i satisfies the intermediate value property, it has to violate strict monotonicity in all \succeq -non-null events.

When preferences are represented by lexicographic multiple priors utility functions defined in (8), Cautious Nash Equilibrium can also be readily constructed to deliver the prediction that player i plays m . When preferences are represented by the lexicographic expected utility model defined in (7), Cautious Nash Equilibrium collapses to *Lexicographic Nash Equilibrium* in which the belief system *respects preferences* (Blume *et al.* (1991b, Definitions 3 and 4)). As explained in section 2, this equilibrium concept can only predict that player i plays l .

4. REFINEMENT OF DOW AND WERLANG (1994)

Motivated by the Ellsberg Paradox, Dow and Werlang (1994) generalize Nash Equilibrium in two-person normal form games. They assume that players' preferences are represented by the Choquet expected utility model where the associated capacity is convex (Schmeidler (1989)). Any such preference ordering is a member of the multiple priors model. In order to see clearly the choice theoretic meaning behind their equilibrium concept, let me first state it in terms of preferences.⁶

Definition 5. $\{\succeq_i, \succeq_j\}$ is a *Nash Equilibrium under Uncertainty* if there exists $E_i \subseteq S_i$ such that the following conditions are satisfied:

1. S_i is \succeq_j -infinitely more likely than the complement of E_i .
2. $s_i \succeq_i \hat{s}_i \quad \forall s_i \in E_i \quad \forall \hat{s}_i \in S_i$;

and similarly for j .

Say that player i is *rational* (*irrational*) if he chooses (does not choose) a strategy that maximizes utility given his beliefs B_i . Nash Equilibrium under Uncertainty requires the whole strategy space S_i to be \succeq_j -infinitely more likely than the event "player i is irrational" (which is a subset of the complement of E_i).

Dow and Werlang (1994, p.313) explicitly adopt the view that the degree of uncertainty aversion is subjective, as in the single agent setting, rather than reasonably tied to the structure of the game. As a result, their equilibrium concept delivers a continuum of equilibria for every normal form game (see their theorem on p.313). In fact, this point is readily demonstrated by noting that given any two-person normal form game, regardless of its payoff structure, the beliefs profile $\{M(S_j), M(S_i)\}$ constitutes a Nash Equilibrium under Uncertainty. Note that in Example 1, if the beliefs of player i are represented by $M(\{L, R\})$, n will be a best response for player i . Therefore it seems desirable

⁶The equilibrium concept proposed by Dow and Werlang, when stated in terms of the multiple priors model, is as follows: $\{B_i, B_j\}$ is a Nash Equilibrium under Uncertainty if there exists $E_i \subseteq S_i$ such that the following conditions are satisfied:

1. $b_j(E_i) = 1$ for at least one $b_j \in B_j$.
2. $\min_{b_i \in B_i} u_i(s_i, b_i) \geq \min_{b_i \in B_i} u_i(\hat{s}_i, b_i) \quad \forall s_i \in E_i \quad \forall \hat{s}_i \in S_i$;

and similarly for j .

To see that this is equivalent to Definition 5, note that S_i is \succeq_j -infinitely more likely than the complement of E_i if and only if there exists $b_j \in B_j$ such that $b_j(\text{complement of } E_i) = 0$ if and only if there exists $b_j \in B_j$ such that $b_j(E_i) = 1$.

to provide a refinement of their equilibrium concept. The following two examples provide further specific motivation.

Example 2

	s_j^1	s_j^2	s_j^3
s_i^1	3,3	3,2	0,3
s_i^2	2,3	2,2	2,3
s_i^3	3,0	3,2	3,3

The following constitutes a Nash Equilibrium under Uncertainty:

$$B_i = \{b_i \in M(S_j) \mid 0.8 \leq b_i(s_j^1) \leq 1, 0 \leq b_i(s_j^2) \leq 0.2, b_i(s_j^3) = 0\},$$

$$B_j = \{b_j \in M(S_i) \mid 0.8 \leq b_j(s_i^1) \leq 1, 0 \leq b_j(s_i^2) \leq 0.2, b_j(s_i^3) = 0\}.$$

To see this, note that according to the beliefs B_j of player j , S_i is \succeq_j -infinitely more likely than $\{s_i^2, s_i^3\}$ and s_i^1 maximizes the utility of player i given i 's beliefs B_i . In this equilibrium, s_i^2 is not \succeq_j -null and s_i^3 is \succeq_j -null. However, s_i^3 strictly dominates s_i^2 . If player j is not even sure whether player i will play s_i^2 , how can she be so sure that i will never play s_i^3 ? Note that in a Nash Equilibrium, given a strategy s_i that is not \succeq_j -null and another strategy \hat{s}_i that is \succeq_j -null, s_i is at least as good as \hat{s}_i for player i . ■

Example 3

	L	R
U	10,10	2,2
D	2,2	1,1

$\{M(S_j), M(S_i)\}$ is a Nash Equilibrium under Uncertainty. According to the beliefs $M(S_i)$ of player j , U and D are not \succeq_j -null. In fact, they are “equally likely” (U is \succeq_j -more likely than D and vice versa). However, U strictly dominates D . Even though player j does not want to completely rule out the possibility that player i will play D , it seems more reasonable that player j believes it is (infinitely) more likely for player i to play U rather than D . Note that in a Nash Equilibrium, given any two strategies s_i and \hat{s}_i that are not \succeq_j -null (and one strategy is not \succeq_j -infinitely more likely than the other), s_i and \hat{s}_i are equally good for player i . ■

Starting from this point, I assume that \succeq obeys the following axiom:

Axiom. Given any state $\omega \in \Omega$ and any two events A and B , if ω is \succeq -infinitely more likely than A and ω is \succeq -infinitely more likely than B , then ω is \succeq -infinitely more likely than $A \cup B$.

The intuition of this axiom is that, for instance, if the decision maker believes that it is \succeq -infinitely more likely for a die to land on the face labelled “1” rather than on any of its edge, and it is \succeq -infinitely more likely for the die to land on the face labelled “1” rather than on any of its corner, then the decision maker believes that it is infinitely more likely that the die will land on the face labelled “1” rather than on any of its edge or corner. When \succeq is represented by the multiple priors model where Δ is the underlying set of probability measures, \succeq satisfies the axiom if and only if *either* (i) the support of Δ is empty *or* (ii) the support of Δ is non-empty and there exists $p \in \Delta$ such that $p(\text{support of } \Delta) = 1$. Therefore if the support of Δ is nonempty, it is the smallest event E_i which satisfies condition 1 in the definition of Nash Equilibrium under Uncertainty.

Under this axiom, Dow and Werlang’s equilibrium concept can be restated in terms of the multiple priors model as follows:⁷

Definition 6. $\{B_i, B_j\}$ is a *Nash Equilibrium under Uncertainty* if the following conditions are satisfied:

1. If the support of B_j is nonempty, then every $s_i \in \text{support of } B_j$ satisfies

$$\min_{b_i \in B_i} u_i(s_i, b_i) \geq \min_{b_i \in B_i} u_i(\hat{s}_i, b_i) \quad \forall \hat{s}_i \in S_i.$$

2. If the support of B_j is empty, then there exists $E_i \subseteq \text{extended support of } B_j$ with $b_j(E_i) = 1$ for at least one $b_j \in B_j$ such that every $s_i \in E_i$ satisfies

$$\min_{b_i \in B_i} u_i(s_i, b_i) \geq \min_{b_i \in B_i} u_i(\hat{s}_i, b_i) \quad \forall \hat{s}_i \in S_i;$$

and similarly for j .

Definition 6 conveys clearly the two messages delivered by Examples 2 and 3. First, in a Nash Equilibrium under Uncertainty, as long as the support of B_j is nonempty, then any comparison among the strategies that are not in the support of B_j is irrelevant. In the Nash Equilibrium under Uncertainty of the game in Example 2, the support of B_j is s_i^1 which is nonempty. s_i^3 is \succeq_j -null and s_i^2 is not \succeq_j -null, but s_i^2 is allowed to strictly dominate s_i^3 . Second, even though the support of B_j is empty and therefore condition 2 in Definition 6 becomes effective, it is not required that every strategy in the extended support of B_j be equally desirable. In the Nash Equilibrium under

⁷Note that the beliefs of the players in all the Nash Equilibria under Uncertainty constructed in the proof of the theorem in Dow and Werlang (1994, p.313) satisfy this axiom.

Uncertainty of the game in Example 3, the support of B_j is empty. U and D are equally likely and not \succeq_j -null, but U is allowed to strictly dominate D .

A direct comparison between Cautious Nash Equilibrium (Definition 4) and Nash Equilibrium under Uncertainty (Definition 6) confirms that, under the above axiom, the former is a refinement of latter. Cautious Nash Equilibrium always requires comparison among some strategies not in the support of B_j . It also requires every strategy, not in the support of B_j , but in its extended support, to be equally desirable. Both Nash Equilibria under Uncertainty constructed for the games in Examples 2 and 3 violate condition 2 in the definition of Cautious Nash Equilibrium. They are therefore refined away.

The solution concept *Equilibrium in ϵ -ambiguous Beliefs* in Mukerji (1994), when restricted to two person games, is essentially the same as Nash Equilibrium under Uncertainty.⁸ Therefore Cautious Nash Equilibrium refines Equilibrium in ϵ -ambiguous Beliefs in exactly the same manner. Finally, Klibanoff (1993) adopts the lexicographic multiple priors model to represent players' preferences in normal form games and defines *Equilibrium with Uncertainty Aversion*. In his paper, the strategy space of each player is the set of mixed strategies.⁹ Therefore a direct comparison with the equilibrium concept here is not possible. However, it is easy to construct an Equilibrium with Uncertainty Aversion in which n is a best response for player i in Example 1.

5. CONCLUSION

In the case where a decision maker is an expected utility maximizer, an event B is \succeq -null if and only if there exists another event A such that A is \succeq -infinitely more likely than B . This relationship may not hold for more general preferences. This paper examines systematically this flexibility in the context of normal form games. It leads to the following contribution: I provide a generalization of Nash Equilibrium in terms of preferences which possesses a sound choice theoretic basis. It can predict prudent behavior as in Example 1. Relative to the equilibrium notions of Dow and Werlang (1994) and Mukerji (1994), my equilibrium concept enjoys higher predictive power.

In this paper, I confine myself to two person normal form games. Nash Equilibrium in n person normal form games requires the beliefs b_i of each player i be *stochastically independent*: b_i is a product measure on $\times_{j \neq i} S_j$. When specialized to the multiple priors model, Cautious Nash Equilibrium can be readily extended to n person normal form games by adopting, for instance, the notion of

⁸Mukerji (1994) requires the capacities to satisfy some additional properties.

⁹See Lo (1995) for arguments for and against the adoption of mixed rather than pure strategy space for normal form games in which the preferences of the players are represented by the multiple priors model.

stochastically independent beliefs in Gilboa and Schmeidler (1989, p.150-151). Generalization of Definition 3 to n person normal form games would require a notion of stochastically independent preferences. One such notion, which is only applicable to preferences satisfying Savage's sure thing principle, can be found in Blume *et al.* (1991a, p.74). Development of a more general notion is a subject for future research.

REFERENCES

- Anscombe, F. J. and R. Aumann (1963): "A Definition of Subjective Probability," *Annals of Mathematical Statistics*, 34, 199-205.
- Aumann, R. (1992): "Irrationality in Game Theory," *Economic Analysis of Markets and Games*, edited by P. Dasgupta, D. Gale, O. Hart and E. Maskin. Cambridge: M.I.T. Press.
- Aumann, R. and A. Brandenberger (1991): "Epistemic Conditions for Nash Equilibrium," *Econometrica*, forthcoming.
- Blume, L., A. Brandenberger and E. Dekel (1991a): "Lexicographic Probabilities and Choice Under Uncertainty," *Econometrica*, 59, 61-79.
- Blume, L., A. Brandenberger and E. Dekel (1991b): "Lexicographic Probabilities and Equilibrium Refinements," *Econometrica*, 59, 81-98.
- Crawford, V. (1990): "Equilibrium without Independence," *Journal of Economic Theory*, 50, 127-154.
- Dekel, E., Z. Safra and U. Segal (1991): "Existence and Dynamic Consistency of Nash Equilibrium with Non-Expected Utility Preferences," *Journal of Economic Theory*, 55, 229-246.
- Dow, J. and S. Werlang (1994): "Nash Equilibrium under Knightian Uncertainty: Breaking Down Backward Induction," *Journal of Economic Theory*, 64, 305-324.
- Gilboa, I. and D. Schmeidler (1989): "Maxmin Expected Utility with Non-unique Prior," *Journal of Mathematical Economics*, 18, 141-153.
- Klibanoff, P. (1993): "Uncertainty, Decision, and Normal Form Games," *Journal of Economic Theory*, forthcoming.
- Lo, K. C. (1995): "Equilibrium in Beliefs under Uncertainty," Manuscript, University of Toronto (first version: 1994).
- Mukerji, S. (1994): "A Theory of Play for Games in Strategic Form When Rationality is not Common Knowledge," Manuscript, Yale University.
- Savage, L. (1954): *The Foundations of Statistics*. New York: John Wiley.
- Schmeidler, D. (1989): "Subjective Probability and Expected Utility without Additivity," *Econometrica*, 57, 571-581.