Econometrica, Vol. 63, No. 5 (September, 1995), 1161–1180

EPISTEMIC CONDITIONS FOR NASH EQUILIBRIUM

BY ROBERT AUMANN AND ADAM BRANDENBURGER[1]

Sufficient conditions for Nash equilibrium in an n-person game are given in terms of what the players know and believe—about the game, and about each other's rationality, actions, knowledge, and beliefs. Mixed strategies are treated not as conscious randomizations, but as conjectures, on the part of other players, as to what a player will do. Common knowledge plays a smaller role in characterizing Nash equilibrium than had been supposed. When $n = 2$, mutual knowledge of the payoff functions, of rationality, and of the conjectures implies that the conjectures form a Nash equilibrium. When $n \geq 3$ and there is a common prior, mutual knowledge of the payoff functions and of rationality, and common knowledge of the conjectures, imply that the conjectures form a Nash equilibrium. Examples show the results to be tight.

KEYWORDS: Game theory, strategic games, equilibrium, Nash equilibrium, strategic equilibrium, knowledge, common knowledge, mutual knowledge, rationality, belief, belief systems, interactive belief systems, common prior, epistemic conditions, conjectures, mixed strategies.

1. INTRODUCTION

IN RECENT YEARS, a literature has emerged that explores noncooperative game theory from a decision-theoretic viewpoint. This literature analyzes games in terms of the rationality[2] of the players and their *epistemic* state: what they know or believe about the game and about each other's rationality, actions, knowledge, and beliefs. As far as Nash's fundamental notion of strategic equilibrium is concerned, the picture remains incomplete;[3] it is not clear just what epistemic conditions lead to Nash equilibrium. Here we aim to fill that gap. Specifically, we seek sufficient epistemic conditions for Nash equilibrium that are in a sense as "spare" as possible.

The stage is set by the following *Preliminary Observation*: *Suppose that each player is rational, knows his own payoff function, and knows the strategy choices of the others. Then the players' choices constitute a Nash equilibrium in the game being played.*[4]

Indeed, since each player knows the choices of the others, and is rational, his choice must be optimal given theirs; so by definition,[5] we are at a Nash equilibrium.

Though simple, the observation is not without interest. Note that it calls for *mutual* knowledge of the strategy choices—that each player know the choices of the others, with no need for the others to know that he knows (or for any higher

[1] We are grateful to Kenneth Arrow, John Geanakoplos, and Ben Polak for important discussions, and to a co-editor and the referees for very helpful editorial suggestions.

[2] We call a player *rational* if he maximizes his utility given his beliefs.

[3] See Section 7i.

[4] For a formal statement, see Section 4.

[5] Recall that a Nash equilibrium is a profile of strategies in which each player's strategy is optimal for him, given the strategies of the others.

order knowledge). It does *not* call for *common* knowledge, which requires that all know, all know that all know, and so on ad infinitum (Lewis (1969)). For rationality and for the payoff functions, not even mutual knowledge is needed; only that the players are in fact rational, and that each knows his own payoff function.[6]

The observation applies to pure strategies—henceforth called *actions*. It applies also to mixed actions, under the traditional view of mixtures as conscious randomizations; in that case it is the mixtures that must be mutually known, not their pure realizations.

In recent years, a different view of mixing has emerged.[7] According to this view, players do not randomize; each player chooses some definite action. But other players need not know which one, and the mixture represents their uncertainty, their conjecture about his choice. This is the context of our main results, which provide sufficient conditions for a profile of conjectures to constitute a Nash equilibrium.[8]

Consider first the case of two players. Here the conjecture of each is a probability distribution on the other's actions—formally, a mixed action of the other. We then have the following (Theorem A): *Suppose that the game being played* (*i.e., both payoff functions*), *the rationality of the players, and their conjectures are all mutually known. Then the conjectures constitute a Nash equilibrium.*[9]

In Theorem A, as in the preliminary observation, common knowledge plays no role. This is worth noting, in view of suggestions that have been made that there is a close relation between Nash equilibrium and common knowledge—of the game, the players' rationality, their beliefs, and/or their choices.[10] On the face

[6] Knowledge of one's own payoff function may be considered tautologous. See Section 2.

[7] Harsanyi (1973), Armbruster and Boege (1979), Aumann (1987a), Tan and Werlang (1988), Brandenburger and Dekel (1989), among others.

[8] The preliminary observation, too, may be interpreted as referring to an equilibrium in conjectures rather than actions. When each player knows the actions of the others, then conjectures coincide with actions: what people do is the same as what others believe them to do. Therefore, the conjectures as well as the actions are in equilibrium.

[9] The idea of the proof is not difficult. Call the players "Rowena" and "Colin"; let their conjectures and payoff functions be ϕ, g and ψ, h respectively. Let a be an action of Rowena to which Colin's conjecture ψ assigns positive probability. Since Colin knows that Rowena is rational, he knows that a is optimal against her conjecture, which he knows to be ϕ, given her payoff function, which he knows to be g. Similarly any action b to which ϕ assigns positive probability is optimal against ψ given Colin's payoff function h. So (ψ, ϕ) is a Nash equilibrium in the game defined by (g, h).

[10] Thus Kreps and Wilson (1982, p. 885): "An equilibrium in Nash's sense supposes that strategies are 'common knowledge' among the players." Or Geanakoplos, Pearce, and Stacchetti (1989, p. 62): "In traditional equilibrium analysis, the equilibrium strategy profile is taken to be common knowledge." Or Milgrom and Roberts (1991, p. 82): "Equilibrium analysis dominates the study of games of strategy, but ... many ... are troubled by its assumption that players ... identify and play a particular vector of equilibrium strategies, that is, ... that the equilibrium is common knowledge." See also, inter alia, Arrow (1986, p. S392), Binmore and Dasgupta (1986, pp. 2–5), Fudenberg and Kreps (1988, p. 2), Tan and Werlang (1988, pp. 381–385), Fudenberg and Tirole (1989, p. 267), Werlang (1989, p. 82), Binmore (1990, pp. 51, 61, 210), Rubinstein (1991, p. 915), Binmore (1992, p. 484), and Reny (1992, p. 628).

We ourselves have written in this vein (Aumann (1987b, p. 473) and Binmore and Brandenburger (1990, p. 119)); see Section 7f.

of it, such a relation sounds not implausible. One might have reasoned that each player plays his part of the equilibrium "because" the other does so; he, in turn, also does so "because" the first does so; and so on ad infinitum. This infinite regress does sound related to common knowledge; but the connection, if any, is murky.[11] Be that as it may, Theorem A shows that in two-person games, epistemic conditions not involving common knowledge in any way already imply Nash equilibrium.

When the number n of players exceeds 2, the conjecture of a player i is not a mixed action of another player, but a probability distribution on $(n-1)$-tuples of actions of all the other players. Though not itself a mixed action, i's conjecture does induce a mixed action[12] for each player j other than i; we call this i's *conjecture about* j. However, different players other than j may have different conjectures about j. Since j's component of the putative equilibrium is meant to represent the conjectures of other players i about j, and these may be different for different i, it is not clear how j's component should be defined.

To proceed, we need another definition. The players are said to have a *common prior*[13] if all differences between their probability assessments are due only to differences in their information; more precisely, if one can think of the situation as arising from one in which the players had the same information and probability assessments, and then got different information.

Theorem B, our n-person result, is now as follows: *In an n-player game, suppose that the players have a common prior, that their payoff functions and their rationality are mutually known, and that their conjectures are commonly known. Then for each player j, all the other players i agree on the same conjecture σ_j about j; and the resulting profile $(\sigma_1, \dots, \sigma_n)$ of mixed actions is a Nash equilibrium.*

So common knowledge enters the picture after all, but in an unexpected way, and only when there are at least three players. Even then, what is needed is common knowledge of the players' conjectures, not of the game or of the players' rationality.

Theorems A and B are formally stated and proved in Section 4.

In the observation as well as the two results, the conditions are sufficient, not necessary. It is always possible for the players to blunder into a Nash equilibrium "by accident," so to speak, without anybody knowing much of anything. Nevertheless, the statements are "tight," in the sense that they cannot be improved upon; none of the conditions can be left out, or even significantly weakened. This is shown by a series of examples in Section 5, which, in addition, provide insight into the role played by the epistemic conditions.

One might suppose that one needs stronger hypotheses in Theorem B than in Theorem A only because when $n \geq 3$, the conjectures of two players about a

[11] Brandenburger and Dekel (1989) do state a relation between common knowledge and Nash equilibrium in the two-person case, but it is quite different from the simple sufficient conditions established here. See Section 7i.

[12] The marginal on j's actions of i's overall conjecture.

[13] Aumann (1987a); for a formal definition, see Section 2. Harsanyi (1967–68) uses the term "consistency" to describe this situation.

third one may disagree. But that is not so. One of the examples in Section 5 shows that even when the necessary agreement is assumed outright, conditions similar to those of Theorem A do not suffice for Nash equilibrium when $n \geq 3$.

Summing up: With two players, mutual knowledge of the game, of the players' rationality, and of their conjectures implies that the conjectures constitute a Nash equilibrium. To reach the same conclusion when there are at least three players, one must also assume a common prior and common knowledge of the conjectures.

The above presentation, while correct, has been informal, and sometimes slightly ambiguous. For an unambiguous presentation, one needs a formal framework for discussing epistemic matters in game contexts; in which, for example, one can describe a situation where each player maximizes against the choices of the others, all know this, but not all know that all know this. In Section 2 we describe such a framework, called an *interactive belief system*; it is illustrated in Section 3. Section 6 defines infinite belief systems, and shows that our results apply to this case as well.

The paper concludes with Section 7, where we discuss conceptual matters and related work.

The reader wishing to understand just the main ideas should read Sections 1 and 5, and skim Sections 2 and 3.

2. INTERACTIVE BELIEF SYSTEMS

Let us be given a strategic *game form*; that is, a finite set $\{1,\dots,n\}$ (the *players*), together with an *action set* A_i for each player i. Set $A := A_1 \times \cdots \times A_n$. An *interactive belief system* (or simply *belief system*) for this game form is defined to consist of:

(2.1) for each player i, a set S_i (i's *types*),

and for each type s_i of i,

(2.2) a probability distribution on the set S^{-i} of $(n-1)$-tuples of types of the other players (s_i's *theory*),

(2.3) an action a_i of i (s_i's *action*), and

(2.4) a function $g_i: A \to \mathbf{R}$ (s_i's *payoff function*).

The action sets A_i are assumed finite. One may also think of the type spaces S_i as finite throughout the paper; the ideas are then more transparent. For a general definition, where the S_i are measurable spaces and the theories are probability measures, see Section 6.

Set $S := S_1 \times \cdots \times S_n$. Call the members $s = (s_1,\dots,s_n)$ of S *states of the world*, or simply *states*. An *event* is a subset E of S. By (2.2), s_i's theory has domain S^{-i}; define an extension $p(\cdot;\, s_i)$ of the theory to S, called i's *probability distribution on S at s_i*, as follows: If E is an event, define $p(E;\, s_i)$ as the probability that s_i's theory assigns to $\{s^{-i} \in S^{-i}\colon (s_i,\, s^{-i}) \in E\}$. Abusing our terminology a little, we will use "belief system" to refer also to the system consisting of S and the $p(\cdot;\, s_i)$; no confusion should result.[14]

A state is a formal description of the players' actions, payoff functions, and beliefs—about each other's actions and payoff functions, about these beliefs, and so on. Specifically, the theory of a type s_i represents the probabilities that s_i ascribes to the types of the other players, and so to their actions, their payoff functions, and their theories. It follows that a player's type determines his beliefs about the actions and payoff functions of the others, about their beliefs about these matters, about their beliefs about others' beliefs about these matters, and so on ad infinitum. The whole infinite hierarchy of beliefs about beliefs about beliefs...about the relevant variables is thus encoded in the belief system.[15]

A function $g\colon A \to \mathbf{R}^n$ (an n-tuple of payoff functions) is called a *game*.

Set $A^{-i} := A_1 \times \cdots \times A_{i-1} \times A_{i+1} \times \cdots \times A_n$; for a in A, set $a^{-i} := (a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_n)$. When referring to a player i, the phrase "at s" means "at s_i." Thus "i's action at s" means s_i's action (see (2.3)); we denote it $\mathbf{a}_i(s)$, and write $\mathbf{a}(s)$ for the n-tuple $(\mathbf{a}_1(s), \dots, \mathbf{a}_n(s))$ of actions at s. Similarly, "i's payoff function at s" means s_i's payoff function (see (2.4)); we denote it $\mathbf{g}_i(s)$, and write $\mathbf{g}(s)$ for the n-tuple $(\mathbf{g}_1(s), \dots, \mathbf{g}_n(s))$ of payoff functions[16] at s. Viewed as a function of a, we call $\mathbf{g}(s)$ "the game being played at s," or simply "the game at s."

Functions defined on S (like $\mathbf{a}_i$, $\mathbf{a}$, $\mathbf{g}_i$, and $\mathbf{g}$) may be viewed like random variables in probability theory. Thus if $\mathbf{x}$ is such a function and x is one of its values, then $[\mathbf{x} = x]$, or simply $[x]$, denotes the event $\{s \in S\colon \mathbf{x}(s) = x\}$. For example, $[a_i]$ denotes the event that i chooses the action a_i, $[g]$ denotes the event that the game g is being played; and $[s_i]$ denotes the event that i's type is s_i.

A *conjecture* ϕ^i of i is a probability distribution on A^{-i}. For $j \neq i$, the marginal of ϕ^i on A_j is called the *conjecture of i about j induced by* ϕ^i. The theory of i at a state s yields a conjecture $\boldsymbol{\phi}^i(s)$, called *i's conjecture at* s, given by $\boldsymbol{\phi}^i(s)(a^{-i}) := p([a^{-i}];\, s_i)$. We denote the n-tuple $(\boldsymbol{\phi}^1(s), \dots, \boldsymbol{\phi}^n(s))$ of conjectures at s by $\boldsymbol{\phi}(s)$.

Player i is called *rational at* s if his action at s maximizes his expected payoff given his information (i.e., his type s_i); formally, letting $g_i := \mathbf{g}_i(s)$ and $a_i := \mathbf{a}_i(s)$, this means that $\exp(g_i(a_i, \mathbf{a}^{-i});\, s_i) \geq \exp(g_i(b_i, \mathbf{a}^{-i});\, s_i)$ for all b_i in A_i. Another way of saying this is that i's actual choice a_i maximizes the expectation of his actual payoff g_i when the other players' actions are distributed according to his actual conjecture $\boldsymbol{\phi}^i(s)$.

[14] The extension $p(\cdot;\, s_i)$ is uniquely determined by two conditions: first, that its marginal on S^{-i} be s_i's theory; second, that it assign probability 1 to i being of type s_i. We are thus implicitly assuming that a player of type s_i assigns probability 1 to being of type s_i. For a discussion, see Section 7c.

[15] Conversely, it may be shown that any such hierarchy satisfying certain minimal coherency requirements may be encoded in some belief system (Mertens and Zamir (1985); also Armbruster and Boege (1979), Boege and Eisele (1979), and Brandenburger and Dekel (1993)).

[16] Thus i's actual payoff at the state s is $\mathbf{g}_i(s)(\mathbf{a}(s))$.

Player i is said to *know* an event E *at* s if at s, he ascribes probability 1 to E. Define K_iE as the set of all those s at which i knows E. Set $K^1E := K_1E \cap \ldots \cap K_nE$; thus K^1E is the event that all players know E. If $s \in K^1E$, call E *mutually known at* s. Set $CKE := K^1E \cap K^1K^1E \cap K^1K^1K^1E \cap \ldots$; if $s \in CKE$, call E *commonly known at* s.

A probability distribution P on S is called a *common prior* if for all players i and all of their types s_i, the conditional distribution of P given s_i is $p(\cdot;\, s_i)$; this implies that for all i, all events E and F, and all numbers π,

(2.5) if $p(E;\, s_i) = \pi p(F;\, s_i)$ for all $s_i \in S_i$, then $P(E) = \pi P(F)$.

In words, (2.5) says that for each player i, if two events have proportional probabilities given any s_i, then they have proportional prior probabilities.[17]

Belief systems provide a formal language for stating epistemic conditions. When we say that a player knows some event E, or is rational, or has a certain conjecture ϕ^i or payoff function g_i, we mean that that is the case at some specific state s of the world. Some of these ideas are illustrated in Section 3.

We end this section with a lemma that is needed in the sequel.

LEMMA 2.6: *Player i knows that he attributes probability π to an event E if and only if he indeed attributes probability π to E.*

PROOF: *If*: Let F be the event that i attributes probability π to E; that is, $F := \{t \in S\colon p(E;\, t_i) = \pi\}$. Thus $s \in F$ if and only if $p(E;\, s_i) = \pi$. Therefore if $s \in F$, then all states u with $u_i = s_i$ are in F, and so $p(F;\, s_i) = 1$; that is, i knows F at s.

Only if: Suppose that i attributes probability $\rho \neq \pi$ to E. By the "if" part of the proof, he must know this, contrary to his knowing that he attributes probability π to E. *Q.E.D.*

3. AN ILLUSTRATION

Consider a belief system in which all types of each player i have the same payoff function g_i, namely that depicted in Figure 1. Thus the game being played is commonly known. Call the row and column players (Players 1 and 2) "Rowena" and "Colin" respectively. The theories are depicted in Figure 2; here C_1 denotes a type of Rowena whose action is C, whereas D_1 and D_2 denote two different types of Rowena whose actions are D. Similarly for Colin. Each square denotes a state, i.e., a pair of types. The two entries in each square denote the probabilities that the corresponding types of Rowena and Colin ascribe to that state. For example, Colin's type d_2 attributes $\frac{1}{2}$-$\frac{1}{2}$ probabilities to Rowena's type being D_1 or D_2. So at the state (D_2, d_2), he knows that Rowena will choose the action D. Similarly, Rowena knows at (D_2, d_2) that Colin will choose d. Since d and D are optimal against each other, both players are rational at (D_2, d_2) and (D, d) is a Nash equilibrium.

[17] Note for specialists: We do not use "mutual absolute continuity."

	c	d
C	2,2	0,0
D	0,0	1,1

FIGURE 1

	c_1	d_1	d_2
C_1	$\frac{1}{2}, \frac{1}{2}$	$\frac{1}{2}, \frac{1}{2}$	0,0
D_1	$\frac{1}{2}, \frac{1}{2}$	0,0	$\frac{1}{2}, \frac{1}{2}$
D_2	0,0	$\frac{1}{2}, \frac{1}{2}$	$\frac{1}{2}, \frac{1}{2}$

FIGURE 2

We have here a typical instance of the preliminary observation. At (D_2, d_2), there is mutual knowledge of the actions D and d, and both players are in fact rational. But the actions are not common knowledge. Though Colin knows that Rowena will play D, she doesn't know that he knows this; indeed, she attributes probability $\frac{1}{2}$ to his attributing probability $\frac{1}{2}$ to her playing C. Moreover, though both players are rational at (D_2, d_2), there isn't even mutual knowledge of rationality there. For example, Colin's type d_1 chooses d, with an expected payoff of $\frac{1}{2}$, rather than c, with an expected payoff of 1; thus this type is irrational. At (D_2, d_2), Rowena attributes probability $\frac{1}{2}$ to Colin being of this irrational type.

Note that the players have a common prior, which assigns probability $1/6$ to each of the six boxes not containing 0's. This, however, is not relevant to the above discussion.

4. FORMAL STATEMENTS AND PROOFS OF THE RESULTS

We now state and prove Theorems A and B formally. For more transparent formulations, see Section 1. We also supply, for the record, a precise, unambiguous formulation of the preliminary observation.

PRELIMINARY OBSERVATION: *Let a be an n-tuple of actions. Suppose that at some state s, all players are rational, and it is mutually known that* $\mathbf{a} = a$. *Then a is a Nash equilibrium.*

THEOREM A: *With $n = 2$ (two players), let g be a game, ϕ a pair of conjectures. Suppose that at some state, it is mutually known that* $\mathbf{g} = g$, *that the players are rational, and that* $\boldsymbol{\phi} = \phi$. *Then (ϕ^2, ϕ^1) is a Nash equilibrium of g.*

The proof of Theorem A uses two lemmas.

LEMMA 4.1: *Let ϕ be an n-tuple of conjectures. Suppose that at some state s, it is mutually known that $\boldsymbol{\phi} = \phi$. Then $\boldsymbol{\phi}(s) = \phi$. (In words: if it is mutually known that the conjectures are ϕ, then they are indeed ϕ.)*

PROOF: Follows from Lemma 2.6. *Q.E.D.*

LEMMA 4.2: *Let g be a game, ϕ an n-tuple of conjectures. Suppose that at some state s, it is mutually known that $\mathbf{g} = g$, that the players are rational, and that $\boldsymbol{\phi} = \phi$. Let a_j be an action of a player j to which the conjecture ϕ^i of some other player i assigns positive probability. Then a_j maximizes g_j against*[18] *ϕ^j.*

PROOF: By Lemma 4.1, the conjecture of i at s is ϕ^i. So i attributes positive probability at s to $[a_j]$. Also, i attributes probability 1 at s to each of the three events [j is rational], $[\phi^j]$, and $[g_j]$. When one of four events has positive probability, and the other three each have probability 1, then their intersection is nonempty. So there is a state t at which all four events obtain: j is rational, he chooses a_j, his conjecture is ϕ^j, and his payoff function is g_j. So a_j maximizes g_j against ϕ^j. *Q.E.D.*

PROOF OF THEOREM A: By Lemma 4.2, every action a_1 with positive probability in ϕ^2 is optimal against ϕ^1 in g, and every action a_2 with positive probability in ϕ^1 is optimal against ϕ^2 in g. This implies that (ϕ^2, ϕ^1) is a Nash equilibrium of g. *Q.E.D.*

THEOREM B: *Let g be a game, ϕ an n-tuple of conjectures. Suppose that the players have a common prior, which assigns positive probability to it being mutually known that $\mathbf{g} = g$, mutually known that all players are rational, and commonly known that $\boldsymbol{\phi} = \phi$. Then for each j, all the conjectures ϕ^i of players i other than j induce the same conjecture σ_j about j, and $(\sigma_1, \ldots, \sigma_n)$ is a Nash equilibrium of g.*

The proof requires several more lemmas. Some of these are standard when "knowledge" means absolute certainty, but not quite as well known when it means probability 1 belief, as here.

LEMMA 4.3: *$K_i(E_1 \cap E_2 \cap \ldots) = K_iE_1 \cap K_iE_2 \cap \ldots$ (a player knows each of several events if and only if he knows that they all obtain).*

PROOF: At s, player i ascribes probability 1 to $E_1 \cap E_2 \cap \ldots$ if and only if he ascribes probability 1 to each of $E_1, E_2, \ldots$. *Q.E.D.*

LEMMA 4.4: *$CKE \subset K_iCKE$ (if something is commonly known, then each player knows that it is commonly known).*

[18] That is, $\exp g_j(a_j, a^{-j}) \geq \exp g_j(b_j, a^{-j})$ for all b_j in A_j, when a^{-j} is distributed according to ϕ^j.

PROOF: Since $K_iK^1F \supset K^1K^1F$ for all F, Lemma 4.3 yields $K_iCKE = K_i(K^1E \cap K^1K^1E \cap \ldots) = K_iK^1E \cap K_iK^1K^1E \cap \ldots \supset K^1K^1E \cap K^1K^1K^1E \cap \ldots \supset CKE$. *Q.E.D.*

LEMMA 4.5: *Suppose P is a common prior, $K_iH \supset H$, and $p(E;\, s_i) = \pi$ for all $s \in H$. Then $P(E \cap H) = \pi P(H)$.*

PROOF: Let H_i be the projection of H on S_i. From $K_iH \supset H$ it follows that $p(H;\, s_i) = 1$ or 0 according as to whether s_i is or is not[19] in H_i. So when $s_i \in H_i$, then $p(E \cap H;\, s_i) = p(E;\, s_i) = \pi = \pi p(H;\, s_i)$; and when $s_i \notin H_i$, then $p(E \cap H;\, s_i) = 0 = \pi p(H;\, s_i)$. The lemma now follows from (2.5). *Q.E.D.*

LEMMA 4.6: *Let Q be a probability distribution on A with[20] $Q(a) = Q(a_i)Q(a^{-i})$ for all a in A and all i. Then $Q(a) = Q(a_1)\ldots Q(a_n)$ for all a.*

PROOF: By induction. For $n = 1$ and 2 the result is immediate. Suppose it is true for $n-1$. From $Q(a) = Q(a_1)Q(a^{-1})$ we obtain, by summing over a_n, that $Q(a^{-n}) = Q(a_1)Q(a_2, \ldots, a_{n-1})$. Similarly $Q(a^{-n}) = Q(a_i)Q(a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_{n-1})$ whenever $i < n$. So the induction hypothesis yields $Q(a^{-n}) = Q(a_1)Q(a_2)\cdots Q(a_{n-1})$. Hence $Q(a) = Q(a^{-n})Q(a_n) = Q(a_1)Q(a_2)\cdots Q(a_n)$. *Q.E.D.*

PROOF OF THEOREM B: Set $F := CK[\phi]$, and let P be the common prior. By assumption, $P(F) > 0$. Set $Q(a) := P([a]\|F)$. We show that for all a and i,

$$Q(a) = Q(a_i)Q(a^{-i}). \tag{4.7}$$

Set $H := [a_i] \cap F$. By Lemmas 4.3 and 4.4, $K_iH \supset H$, since i knows his own action. If $s \in H$, it is commonly, and so mutually, known at s that $\boldsymbol{\phi} = \phi$; so by Lemma 4.1, $\boldsymbol{\phi}(s) = \phi$; that is, $p([a^{-i}];\, s_i) = \phi^i(a^{-i})$. So Lemma 4.5 (with $E = [a^{-i}]$) yields $P([a] \cap F) = P([a^{-i}] \cap H) = \phi^i(a^{-i})P(H) = \phi^i(a^{-i})P([a_i] \cap F)$. Dividing by $P(F)$ yields $Q(a) = \phi^i(a^{-i})Q(a_i)$; then summing over a_i, we get

$$Q(a^{-i}) = \phi^i(a^{-i}). \tag{4.8}$$

Thus $Q(a) = Q(a^{-i})Q(a_i)$, which is (4.7).

For each j, define a probability distribution σ_j on A_j by $\sigma_j(a_j) := Q(a_j)$. Then (4.8) yields $\phi^i(a_j) = Q(a_j) = \sigma_j(a_j)$ for $j \neq i$. Thus for all i, the conjecture about j induced by ϕ^i is σ_j, which does not depend on i. Lemma 4.6, (4.7), and (4.8) then yield

$$\phi^i(a^{-i}) = \sigma_1(a_1)\cdots\sigma_{i-1}(a_{i-1})\sigma_{i+1}(a_{i+1})\cdots\sigma_n(a_n); \tag{4.9}$$

that is, the distribution ϕ^i is the product of the distributions σ_j with $j \neq i$.

[19] In particular, i always knows whether or not H obtains.

[20] We denote $Q(a^{-i}) := Q(A_i \times \{a^{-i}\})$, $Q(a_i) := Q(A^{-i} \times \{a_i\})$, and so on.

Since common knowledge implies mutual knowledge, the hypothesis of the theorem implies that there is a state at which it is mutually known that $\mathbf{g} = g$, that the players are rational, and that $\boldsymbol{\phi} = \phi$. So by Lemma 4.2, each action a_j with $\phi^i(a_j) > 0$ for some $i \neq j$ maximizes g_j against ϕ^j. By (4.9), these a_j are precisely the ones that appear with positive probability in σ_j. Again using (4.9), we conclude that each action appearing with positive probability in σ_j maximizes g_j against the product of the distributions σ_k with $k \neq j$. This implies that $(\sigma_1, \ldots, \sigma_n)$ is a Nash equilibrium of g. *Q.E.D.*

5. TIGHTNESS OF THE RESULTS

This section explores possible variations on Theorem B. For simplicity, let $n = 3$ (three players). Each player's "overall" conjecture is then a distribution on pairs of actions of the other two players; so the three conjectures form a triple of probability mixtures of action pairs. On the other hand, an equilibrium is a triple of mixed actions. Our discussion hinges on the relation between these two kinds of objects.

First, since our real concern is with mixtures of *actions* rather than of action *pairs*, could we not formulate conditions that deal directly with each player's "individual" conjectures—his conjectures about each of the other players—rather than with his overall conjecture? For example, one might hope that it would be sufficient to assume common knowledge of each player's individual conjectures.

Example 5.1 shows that this hope is vain, even when there is a common prior, and rationality and payoff functions are commonly known. Overall conjectures do play an essential role.

Nevertheless, *common* knowledge of the overall conjectures seems a rather strong assumption. Couldn't we get away with less—say, with mutual knowledge of the overall conjectures, or with mutual knowledge to a high order?[21]

Again, the answer is no. Recall that people with a common prior but with different information may disagree on their posterior probabilities for some event E, even though these posteriors are mutually known to an arbitrarily high order (Geanakoplos and Polemarchakis (1982)). Using this, one may construct an example with arbitrarily high order mutual knowledge of the overall conjectures, common knowledge of rationality and payoff functions, and a common prior, where different players have different individual conjectures about some particular player j. Thus there isn't even a clear *candidate* for a Nash equilibrium.[22]

The question remains whether (sufficiently high order) mutual knowledge of the overall conjectures implies Nash equilibrium of the individual conjectures when the players do happen to agree on them. Do we then get Nash equilibrium? Again, the answer is no; this is shown in Example 5.2.

[21] Set $K^2E := K^1K^1E$, $K^3E := K^1K^2E$, and so on. If $s \in K^mE$, call E *mutually known to order* m at s.

[22] This is *not* what drives Example 5.1; since the individual conjectures are commonly known there, they must agree (Aumann (1976)).

Finally, Example 5.3 shows that the common prior assumption is really needed: Rationality, payoff functions, and the overall conjectures are commonly known, and the individual conjectures agree; but there is no common prior, and the agreed-upon individual conjectures do *not* form a Nash equilibrium.

Summing up, one *must* consider the overall conjectures; and nothing less than common knowledge of these conjectures, together with a common prior, will do.

Also, one may construct examples showing that in Theorems A and B, mutual knowledge of rationality cannot be replaced by the simple fact of rationality, and that knowing one's own payoff function does not suffice—all payoff functions must be mutually known.

Except in Example 5.3, the belief systems in this section have common priors, and these are used to describe them. In all the examples, the game being played is (as in Section 3) fixed throughout the belief system, and so is commonly known. Each example has three players, Rowena, Colin, and Matt, who choose the row, column, and matrix (west or east) respectively. As in Section 3, each type is denoted by the same letter as its action, and a subscript is added.

EXAMPLE 5.1: Here the individual conjectures are commonly known and agreed upon, rationality is commonly known, and there is a common prior, and yet we don't get Nash equilibrium.[23] Consider the game of Figure 3, with theories induced by the common prior in Figure 4. At each state, Colin and Matt agree on the conjecture $\frac{1}{2}U + \frac{1}{2}D$ about Rowena, and this is commonly known. Similarly, it is commonly known that Rowena and Matt agree on the conjecture $\frac{1}{2}L + \frac{1}{2}R$ about Colin, and that Rowena and Colin agree on $\frac{1}{2}W + \frac{1}{2}E$ about Matt. All players are rational at all states, so rationality is common knowledge at all states. But $(\frac{1}{2}U + \frac{1}{2}D,\ \frac{1}{2}L + \frac{1}{2}R,\ \frac{1}{2}W + \frac{1}{2}E)$ is not a Nash equilibrium, because if these were independent mixed strategies, Rowena could gain by moving to D.

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

W

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

E

FIGURE 3

	L_1	R_1
U_1	$\frac{1}{4}$	0
D_1	0	$\frac{1}{4}$

W_1

	L_1	R_1
U_1	0	$\frac{1}{4}$
D_1	$\frac{1}{4}$	0

E_1

FIGURE 4

[23] Examples 2.5, 2.6, and 2.7 of Aumann (1974) display correlated equilibria that are not Nash, but they are quite different from Example 5.1. First, the context there is *global*, as opposed to the *local* context considered here (Section 7i). Second, even if we do adapt those examples to the local context, we find that the individual conjectures are not even mutually known, to say nothing of being commonly known; and when there are more than two players (Examples 2.5 and 2.6), the individual conjectures are not agreed upon either.

Note that the overall conjectures are not commonly (nor even mutually) known at any state. For example, at (U_1, L_1, W_1), Rowena's conjecture is $(\frac{1}{2}LW + \frac{1}{2}RE)$, but nobody else knows that that is her conjecture.

EXAMPLE 5.2: Here we have mutual knowledge of the overall conjectures, agreement of individual conjectures, common knowledge of rationality, and a common prior, and yet the individual conjectures do not form a Nash equilibrium. Consider the game of Figure 5. For Rowena and Colin, this is simply "matching pennies;" their payoffs are not affected by Matt's choice. So at a Nash equilibrium, they must play $\frac{1}{2}H + \frac{1}{2}T$ and $\frac{1}{2}h + \frac{1}{2}t$ respectively. Thus Matt's expected payoff is $\frac{3}{2}$ for W, and 2 for E; so he must play E. Hence $(\frac{1}{2}H + \frac{1}{2}T, \frac{1}{2}h + \frac{1}{2}t, E)$ is the unique Nash equilibrium of this game.

Consider now the theories induced by the common prior in Figure 6. Rowena and Colin know which of the three large boxes contains the true state, and in fact this is commonly known between the two of them. In each box, Rowena and Colin "play matching pennies optimally;" their conjectures about each other are $\frac{1}{2}H + \frac{1}{2}T$ and $\frac{1}{2}h + \frac{1}{2}t$. Since these conjectures obtain at each state, they are commonly known (among all three players); so it is also commonly known that Rowena and Colin are rational.

As for Matt, suppose first that he is of type W_1 or W_2. Each of these types intersects two adjacent boxes in Figure 6; it consists of the diagonal states in the left box and the off-diagonal ones in the right box. The diagonal states on the left have equal probability, as do the off-diagonal ones on the right; but on the left it is three times on the right. So Matt assigns the diagonal states three times the probability of the off-diagonal states; i.e., his conjecture is $\frac{3}{8}Hh + \frac{3}{8}Tt + \frac{1}{8}Th + \frac{1}{8}Ht$. Therefore his expected payoff from choosing W is $\frac{3}{8}\cdot 3 + \frac{3}{8}\cdot 3 + \frac{1}{8}\cdot 0 + \frac{1}{8}\cdot 0 = 2\frac{1}{4}$, whereas from E it is only 2 (as all his payoffs in the eastern matrix are 2). So W is indeed the optimal action of these types; so they are

	h	t
H	1, 0, 3	0, 1, 0
T	0, 1, 0	1, 0, 3

W

	h	t
H	1, 0, 2	0, 1, 2
T	0, 1, 2	1, 0, 2

E

FIGURE 5

	h_1	t_1	h_2	t_2	h_3	t_3
H_1	$9x$ W_1	$9x$ E_1				
T_1	$9x$ E_1	$9x$ W_1			$(x = 1/52)$	
H_2			$3x$ W_2	$3x$ W_1		
T_2			$3x$ W_1	$3x$ W_2		
H_3					x W_3	x W_2
T_3					x W_2	x W_3

FIGURE 6

	h_1	t_1
H_1	$\frac{1}{2}, \frac{1}{2}, \frac{3}{8}$	$\frac{1}{2}, \frac{1}{2}, \frac{1}{8}$
T_1	$\frac{1}{2}, \frac{1}{2}, \frac{1}{8}$	$\frac{1}{2}, \frac{1}{2}, \frac{3}{8}$

W_1

FIGURE 7

rational. It may be checked that also E_1 and W_3 are rational. Thus the rationality of all players is commonly known at all states.

Consider now the state $s := (H_2, h_2, W_2)$ (the top left state in the middle box). Rowena and Colin know at s that they are in the middle box, so they know that Matt's type is W_1 or W_2. We have just seen that these two types have the same conjecture, so it follows that Matt's conjecture is mutually known at s. Also Rowena's and Colin's conjectures are mutually known at s (Rowena's is $\frac{1}{2}hW + \frac{1}{2}tW$, Colin's is $\frac{1}{2}HW + \frac{1}{2}TW$).

Finally, the individual conjectures derived from Matt's overall conjecture $\frac{3}{8}Hh + \frac{3}{8}Tt + \frac{1}{8}Th + \frac{1}{8}Ht$ are $\frac{1}{2}H + \frac{1}{2}T$ for Rowena and $\frac{1}{2}h + \frac{1}{2}t$ for Colin. These are the same as Rowena's and Colin's conjectures about each other. Since Matt plays W throughout the middle box, both Rowena and Colin conjecture W for Colin there. Thus throughout the middle box, individual conjectures are agreed upon.

To sum up: There is a common prior; at all states, the game is commonly known and all players are commonly known to be rational. At the top left state in the middle box, the overall conjectures of all players are mutually known, and the individual conjectures are agreed: $\sigma_R = \frac{1}{2}H + \frac{1}{2}T$, $\sigma_C = \frac{1}{2}h + \frac{1}{2}t$, $\sigma_M = W$. But $(\sigma_R, \sigma_C, \sigma_M)$ is not a Nash equilibrium.

One can construct similar examples in which the mutual knowledge of the conjectures is of arbitrarily high order, simply by using more boxes; the result follows as before.

EXAMPLE 5.3: Here we show that one cannot dispense with common priors in Theorem B. Consider again the game of Figure 5, with the theories depicted in Figure 7 (presented in the style of Figure 2). At each state there is common knowledge of rationality, of the overall conjectures (which are the same as in the previous example), and of the game. As before, individual conjectures are in agreement. And as before, the individual conjectures $(\frac{1}{2}H + \frac{1}{2}T, \frac{1}{2}h + \frac{1}{2}t, W)$ do not constitute a Nash equilibrium.

6. GENERAL (INFINITE) BELIEF SYSTEMS

For a general definition of a belief system, which allows it to be infinite,[24] we specify that the type spaces S_i be measurable spaces. As before, a *theory* is a

[24] Infinite belief systems are essential for a complete treatment: The belief system required to encode a given coherent hierarchy of beliefs (see footnote 15) is often uncountably infinite.

probability measure on $S^{-i} = \times_{j \neq i} S_j$, which is now endowed with the standard product structure.[25] The state space $S = \times_j S_j$, too, is endowed with the product structure. An *event* is now a measurable subset of S. The "action functions" $\mathbf{a}_i$ ((2.3)) are assumed measurable; so are the payoff functions $\mathbf{g}_i$ ((2.4)), as functions of s_i, for each action n-tuple a separately. Also the "theory functions" (2.2) are assumed measurable, in the sense that for each event E and player i, the probability $p(E; s_i)$ is measurable as a function of the type s_i. It follows that the conjectures ϕ^i are also measurable functions of s_i.

With these definitions, the statements of the results make sense, and the proofs remain correct, without any change.

7. DISCUSSION

a. *Belief Systems*. An interactive belief system is not a prescriptive model; it does not suggest actions to the players. Rather, it is a formal framework—a language—for *talking* about actions, payoffs, and beliefs. For example, it enables us to say whether a given player is behaving rationally at a given state, whether this is known to another player, and so on. But it does not prescribe or even suggest rationality; the players do whatever they do. Like the disk operating system of a personal computer, the belief system simply organizes things, so that we can coherently discuss what they do.

Though entirely apt, use of the term "state of the world" to include the actions of the players has perhaps caused confusion. In Savage (1954), the decision maker cannot affect the state; he can only react to it. While convenient in Savage's one-person context, this is not appropriate in the interactive, many-person world under study here. To describe the state of such a world, it is fitting to consider all the players simultaneously; and then since each player must take into account the actions of the others, the actions should be included in the description of the state. Also the plain, everyday meaning of the term "state of the world" includes one's actions: Our world is shaped by what we do.

It has been objected that since the players' actions are determined by the state, they have no freedom of choice. But this is a misunderstanding. Each player may do whatever he wants. It is simply that whatever he does do is part of the description of the state. If he wishes to do something else, he is heartily welcome to it; but then the state is different.

Though including one's own action in the state is not a new idea,[26] it may still leave some readers uncomfortable. Perhaps this discomfort stems from a notion that actions should be part of the solution, whereas including them in the state might suggest that they are part of the problem.

The "problem-solution" viewpoint is the older, classical approach of game theory. The viewpoint adopted here is different—it is *descriptive*. Not *why* the

[25] The σ-field of measurable sets is the smallest σ-field containing all the "rectangles" $\times_{j \neq i} T_j$, where T_j is measurable in S_j.

[26] See, e.g., Aumann (1987a).

players do what they do, not what *should* they do; just what *do* they do, what *do* they believe. Are they rational, are they irrational, are their actions—or beliefs—in equilibrium? Not "why," not "should," just *what*. Not that i does a *because* he believes b; simply that he does a, and believes b.

The idea of belief system is due to John Harsanyi (1967–68), who introduced the concept of *I-game* to enable a coherent formulation of games in which the players needn't know each other's payoff functions. I-games are just like belief systems, except that in I-games a player's type does not determine his action (only his payoff function).

As indicated above, belief systems are primarily a convenient framework to enable *us*—the analysts—to discuss the things we want to discuss: actions, payoffs, beliefs, rationality, equilibrium, and so on. As for the players themselves, it's not clear that they need concern themselves with the structure of the model. But if they, too, want to talk about the things that we want to talk about, that's OK; it's just as convenient a framework for them as it is for us. In this connection, we note that the belief system itself may always be considered common knowledge among the players. Formally, this follows from the work of Mertens and Zamir (1985); for an informal discussion, see Aumann (1987a, p. 9 ff.).

b. *Knowledge and Belief*: In this paper, "know" means "ascribe probability 1 to." This is sometimes called "believe," while "know" is reserved for absolute certainty with no possibility at all for error. Since our conditions are sufficient, the results are stronger with probability 1 than with absolute certainty. If probability 1 knowledge of certain events implies that σ is a Nash equilibrium, then a fortiori, so does absolute certainty of those events.

c. *Knowledge of One's Own Type*: It is implicit in our set-up that each player i knows his own type s_i—that is, he knows his theory, his payoff function g_i, and his action a_i.

Knowledge of one's theory is not a substantive restriction; the theory consists of beliefs, and it is tautologous that one knows what one's beliefs are (a formal expression of this is Lemma 2.6).

Knowledge of one's payoff function is a more subtle matter. On the face of it, it would seem quite possible for a player's payoff to depend on circumstances known to others but not to himself. In our set-up, this could be expressed by saying that a player's payoff might depend on the types of other players as well as his own. To avoid this, one may interpret $g_i(a)$ as expressing the payoff that i *expects* when a is played, rather than what he actually gets. And since one always knows one's own expectation, one may as well construct the system so that knowledge of one's own payoff is tautological.

We come finally to knowledge of one's own action. If one thinks of actions as conscious choices, as we do here, this is very natural—one might almost say tautologous. That players are aware of—"know"—their own conscious choices is implicit in the word "conscious."

Of course, if explicit randomization is allowed, then the players need *not* be aware of their own pure actions. But even then, they *are* aware of the mixtures they choose; so *mutatis mutandis*, our analysis applies to the mixtures. See Section 1 for a brief discussion of this case; it is not our main concern here, where we think of i's mixed actions as representing the beliefs of other players about what i will do.

d. *Knowledge of Conjectures*: Both our theorems assume some form of knowledge (mutual or common) of the players' *conjectures*. Though knowledge of what others will *do* is undoubtedly a strong assumption, one can imagine circumstances in which it would obtain. But can one know what others *think*? And if so, can this happen in contexts of economic interest?

In fact, it might happen in several ways. One has to do with players who are members of well-defined economic populations, like insurance companies and customers, or sellers and buyers in general. For example, someone is buying a car. She knows that the salesman has statistical information about customers' bargaining behavior, and she even knows what that statistical information is. So she knows the salesman's conjecture about her. The conjecture may even be commonly known by the two players. But it is more likely that though the customer knows the salesman's conjecture about her, she does not know that he knows that she knows, and indeed perhaps he doesn't; then the knowledge of the salesman's conjecture is only mutual.

No doubt, this story has its pros and cons; we don't want to make too much of it. It is meant only to show that a player may well know another's conjecture in situations of economic interest.

e. *Knowledge of Equilibrium*: Our results state that a specified (mixed) strategy n-tuple σ is an equilibrium; they do not state that the players know it to be an equilibrium, or that this is commonly known. In Theorems A and B, though, it is in fact mutual knowledge to order 1—but not necessarily to any higher order—that σ is a Nash equilibrium. In the preliminary observation, it need not even be mutual knowledge to order 1 that σ is a Nash equilibrium; but this does follow if, in addition to the stated assumptions, one assumes mutual knowledge of the payoff functions.

f. *Common Knowledge of the Model*: Binmore and Brandenburger (1990, p. 119) have written that "in game theory, it is typically understood that the structure of the game...is common knowledge." In the same vein, Aumann (1987b, p. 473) has written that "the common knowledge assumption underlies all of game theory and much of economic theory. Whatever be the model under discussion,...the model itself must be common knowledge; otherwise the model is insufficiently specified, and the analysis incoherent." This seemed sound when written, but in the light of recent developments—including the present work[27]

[27] Specifically, that common knowledge of the payoff functions plays no role in our theorems.

—it no longer does. Admittedly, we do use a belief system, which is, in fact, commonly known. But the belief system is not a "model" in the sense of being exogenously given; it is merely a language for discussing the situation (see Section 7a). There is nothing about the real world that must be commonly known among the players. When writing the above, we thought that some real exogenous framework must be commonly known; this no longer seems appropriate.

g. *Independent Conjectures*: The proof of Theorem B implies that the individual conjectures of each player i about the other players j are independent. Alternatively, one could *assume* independence, as in the following:

REMARK 7.1: Let σ be an n-tuple of mixed strategies. Suppose that at some state, it is mutually known that the players are rational, that the game g is being played, that the conjecture of each player i about each other player j is σ_j, and that it is independent of i's conjecture about all other players. Then σ is a Nash equilibrium in g.

Here we assume mutual rather than common knowledge of conjectures and do not assume a common prior. On the other hand, we assume outright that the individual conjectures are agreed upon, and that each player's conjectures about the others are independent. We consider this result of limited interest in the context of this paper; neither assumption has the epistemic flavor that we are seeking. Moreover, in the current subjectivist context, we find independence dubious as an assumption (though not necessarily as a conclusion). See Aumann (1987a, p. 16).

h. *Converses*: We have already mentioned (at the end of Section 1) that our conditions are not necessary, in the sense that it is quite possible to have a Nash equilibrium even when they are not fulfilled. Nevertheless, there is a sense in which the converses hold: Given a Nash equilibrium in a game g, one can construct a belief system in which the conditions are fulfilled. For the preliminary observation, this is immediate: Choose a belief system where each player i has just one type, whose action is i's component of the equilibrium and whose payoff function is g_i. For Theorems A and B, we may suppose that as in the traditional interpretation of mixed strategies, each player chooses an action by an independent conscious randomization according to his component σ_i of the given equilibrium σ. The types of each player correspond to the different possible outcomes of the randomization; each type chooses a different action. All types of player i have the same theory, namely, the product of the mixed strategies of the other $n-1$ players appearing in σ, and the same payoff function, namely g_i. It may then be verified that the conditions of Theorems A and B are met.

These "converses" show that the sufficient conditions for Nash equilibrium in our theorems are not too strong, in the sense that they do not imply more than Nash equilibrium; every Nash equilibrium is attainable with these conditions. Another sense in which they are not too strong—that the conditions cannot be dispensed with or even appreciably weakened—was discussed in Section 5.

i. *Related Work*: This paper joins a growing epistemic literature in noncooperative game theory. For two-person games, Tan and Werlang (1988) show that if the players' payoff functions, conjectures, and rationality are all commonly known, then the conjectures constitute a Nash equilibrium.[28] Brandenburger and Dekel (1989) take this further. They ask, "when is Nash equilibrium equivalent to common knowledge of rationality? When do these two basic ideas, one from game theory and the other from decision theory, coincide?" The answer they provide is that in two-person games, a sufficient condition for this is that the payoff functions and conjectures be commonly known.[29] That is, *if* the payoff functions and the conjectures are commonly known, *then* rationality is commonly known if and only if the conjectures constitute a Nash equilibrium. The "only if" part of this is precisely the above result of Tan-Werlang.[30, 31]

Our Theorem A improves on the Tan-Werlang result in that it assumes only mutual knowledge where they assume common knowledge.

Aumann (1987a) is also part of the epistemic literature, but the question it addresses is quite different from that of this paper. It asks about the distribution of action profiles over the entire state space when all players are assumed rational at all states of the world and there is a common prior. The answer is that it represents a correlated (not Nash!) equilibrium. Conceptually, that paper takes a *global* point of view; its result concerns the distribution of action profiles as a whole. Correlated equilibrium itself is an essentially global concept; it has no natural local formulation. In contrast, the viewpoint of the current paper is *local*. It concerns the information of the players, at some specific state of the world; and it asks whether the players' actions or conjectures *at that state* constitute a Nash equilibrium. Matters like knowledge that is mutual but not

[28] This is a restatement of their result in our formalism and terminology, which differ substantially from theirs. In particular, two players' "knowing each other" in the sense of Tan and Werlang implies that in our sense, their conjectures are common knowledge.

[29] For example, if there is an accepted convention as to how to act in a certain situation (like driving on the right), then that convention constitutes a Nash equilibrium if and only if it is commonly known that everyone is acting rationally.

[30] We may add that Armbruster and Boege (1979) also treat two-person games in this spirit.

[31] One may ask whether our theorems can be extended to equivalence results in the style of Brandenburger-Dekel. The answer is yes. For two-person games, it can be shown that if the payoff functions and the conjectures are mutually known, then rationality is mutually known if and only if the conjectures constitute a Nash equilibrium; this extends Theorem A. Theorem B may be extended in a similar fashion, as may the preliminary observation. The "only if" parts of these extensions coincide with the theorems established in the body of this paper.

The "if" parts of these extensions are interesting, but perhaps not as much as the "if" part of the Brandenburger-Dekel result: Mutual knowledge of rationality is weaker than common knowledge, and so it is less appealing as a conclusion. That is one reason that we have not made more of these extensions.

common, or players who are rational but may ascribe irrationality to one another, do not come under the purview of Aumann (1987a).

Brandenburger and Dekel (1987, Proposition 4.1) derive Nash equilibrium in n-person games from independence assumptions, as we do[32] in Proposition 7.1. As we have already noted (Section 7g), such assumptions lack the epistemic flavor that interests us here.

In brief: The aim of this paper has been to identify sufficient epistemic conditions for Nash equilibrium that are as spare as possible; to isolate just what the players themselves must know in order for their conjectures about each other to constitute a Nash equilibrium. For two-person games, our Theorem A goes significantly beyond that done previously in the literature on this issue. For n-person games, little had been done before.

Institute of Mathematics, The Hebrew University, 91904 Jerusalem, Israel,
and
Harvard Business School, Boston, MA 02163, U.S.A.

Manuscript received September, 1991; final revision received October, 1994.

REFERENCES

ARMBRUSTER, W., AND W. BOEGE (1979): "Bayesian Game Theory," in *Game Theory and Related Topics*, ed. by O. Moeschlin and D. Pallaschke. Amsterdam: North-Holland.

ARROW, K. (1986): "Rationality of Self and Others in an Economic System," *Journal of Business*, 59, S385–S399.

AUMANN, R. (1974): "Subjectivity and Correlation in Randomized Strategies," *Journal of Mathematical Economics*, 1, 67–96.

——— (1976): "Agreeing to Disagree," *Annals of Statistics*, 4, 1236–1239.

——— (1987a): "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica*, 55, 1–18.

——— (1987b): "Game Theory," in *The New Palgrave: A Dictionary of Economics*, ed. by J. Eatwell, M. Milgate, and P. Newman. London: The MacMillan Press Ltd.

BINMORE, K. (1990): *Essays on the Foundations of Game Theory*. Oxford: Basil Blackwell.

——— (1992): *Fun and Games*. Lexington: D.C. Heath and Company.

BINMORE, K., AND A. BRANDENBURGER (1990): "Common Knowledge and Game Theory," in *Essays on the Foundation of Game Theory*, ed. by K. Binmore. Oxford: Basil Blackwell, pp. 105–150.

BINMORE, K., AND P. DASGUPTA (1986): "Game Theory: A Survey," in *Economic Organizations as Games*, ed. by K. Binmore and P. Dasgupta. Oxford: Basil Blackwell, 1–45.

BOEGE, W., AND T. EISELE (1979): "On Solutions of Bayesian Games," *International Journal of Game Theory*, 8, 193–215.

BRANDENBURGER, A., AND E. DEKEL (1987): "Rationalizability and Correlated Equilibria," *Econometrica*, 55, 1391–1402.

——— (1989): "The Role of Common Knowledge Assumptions in Game Theory," in *The Economics of Missing Markets, Information, and Games*, ed. by F. Hahn. Oxford: Oxford University Press, 46–61.

——— (1993): "Hierarchies of Beliefs and Common Knowledge," *Journal of Economic Theory*, 59, 189–198.

[32] Though their result (henceforth B&D4) differs from our Proposition 7.1 (hereforth P7) in several ways. First, B&D4 makes an assumption tantamount to common knowledge of conjectures, while P7 asks only for mutual knowledge. Second, P7 directly assumes agreement among the individual conjectures, which B&D4 does not. Finally, B&D4 requires "concordant" priors (a weaker form of common priors), while P7 does not.

FUDENBERG, D., AND D. KREPS (1988): "A Theory of Learning, Experimentation, and Equilibrium in Games," unpublished, Graduate School of Business, Stanford University.

FUDENBERG, D., AND J. TIROLE (1989): "Noncooperative Game Theory for Industrial Organization: An Introduction and Overview," in *Handbook of Industrial Organization*, Vol. 1, ed. by R. Schmalensee and R. Willig. Amsterdam: North-Holland, 259–327.

GEANAKOPLOS, J., D. PEARCE, AND E. STACCHETTI (1989): "Psychological Games and Sequential Rationality," *Games and Economic Behavior*, 1, 60–79.

GEANAKOPLOS, J., AND H. POLEMARCHAKIS (1982): "We Can't Disagree Forever," *Journal of Economic Theory*, 28, 192–200.

HARSANYI, J. (1967–68): "Games with Incomplete Information Played by 'Bayesian' Players," Parts I-III, *Management Science*, 8, 159–182, 320–334, 486–502.

——— (1973): "Games with Randomly Disturbed Payoffs: A New Rationale for Mixed Strategy Equilibrium Points," *International Journal of Game Theory*, 2, 1–23.

KREPS, D., AND R. WILSON (1982): "Sequential Equilibria," *Econometrica*, 50, 863–894.

LEWIS, D. (1969): *Convention: A Philosophical Study*. Cambridge: Harvard University Press.

MERTENS, J-F., AND S. ZAMIR (1985): "Formulation of Bayesian Analysis for Games with Incomplete Information," *International Journal of Game Theory*, 14, 1–29.

MILGROM, P., AND J. ROBERTS (1991): "Adaptive and Sophisticated Learning in Normal Form Games," *Games and Economic Behavior*, 3, 82–100.

NASH, J. (1951): "Non-Cooperative Games," *Annals of Mathematics*, 54, 286–295.

RENY, P. (1992): "Backward Induction, Normal Form Perfection and Explicable Equilibria," *Econometrica*, 60, 627–649.

RUBINSTEIN, A. (1991): "Comments on the Interpretation of Game Theory," *Econometrica*, 59, 909–924.

SAVAGE, L. (1954): *The Foundations of Statistics*. New York: Wiley.

TAN, T., AND S. WERLANG (1988): "The Bayesian Foundations of Solution Concepts of Games," *Journal of Economic Theory*, 45, 370–391.

WERLANG, S. (1989): "Common Knowledge," in *The New Palgrave: Game Theory*, ed. by J. Eatwell, M. Milgate, and P. Newman. New York: W. W. Norton, 74–85.