Epistemic Conditions for Nash Equilibrium*

Robert Aumann† Adam Brandenburger‡

1991 Version

Abstract

According to conventional wisdom, Nash equilibrium in a game "involves" common knowledge of the payoff functions, of the rationality of the players, and of the strategies played. The basis for this wisdom is explored, and it turns out that considerably weaker conditions suffice. First, note that if each player is rational and knows his own payoff function, and the strategy choices of the players are mutually known, then these choices form a Nash equilibrium. The other two results treat the mixed strategies of a player not as conscious randomization of that player, but as conjectures of the other players about what he will do. When $n = 2$, mutual knowledge of the payoff functions, of rationality, and of the conjectures yields Nash equilibrium. When $n \geq 3$, mutual knowledge of the payoff functions and of rationality, and common knowledge of the conjectures yield Nash equilibrium when there is a common prior. Examples are provided showing these results to be sharp.

1 Introduction

Game theoretic reasoning has been widely applied in economics in recent years. Undoubtedly, the most commonly used tool has been the strategic equilibrium of Nash [15, 1951], or one or another of its so-called "refinements." Though much effort[1] has gone into developing these refinements, relatively little attention has been paid to a more basic question: Why consider Nash equilibrium in the first place?

A Nash equilibrium is defined as a way of playing the game–an n-tuple of strategies–in which each player's strategy is optimal for him, given the strategies of the others. The definition seems beautifully simple and compelling; but when considered carefully, it lacks a clear motivation. What would make the players play such an equilibrium? How, exactly, would it come about?

*We are grateful to Kenneth Arrow, John Geanakoplos, and Ben Polak for important input ot this paper. Financial support from NSF grant IRI-8814953 at the Economic Department, Stanford University, and the Harvard Business School Division of Research is gratefully acknowledged. ecne-10-03-06

†Address: Institute of Mathematics, The Hebrew University, 91904 Jerusalem, Israel

‡Address (then): Harvard Business School Boston, MA 02163

[1]Selten [17, 1965], [18, 1975], Myerson [14, 1978], Kreps and Wilson [10, 1982], Kohlberg and Mertens [9, 1986], and many others.

Over the years, much has been written about the connection between Nash equilibrium and common knowledge.[2] According to conventional wisdom, Nash equilibrium is "based on" common knowledge of (a) the structure of the game (i.e., the payoff functions[3]), (b) the rationality[4] of the players, and (c) the strategies actually played. These ideas sound appealing; the circularity of Nash equilibrium–each player chooses his strategy only because the others choose theirs–does seem related to the infinite hierarchy of beliefs inherent in common knowledge. But a formalization has proved elusive. What, precisely, does "based on" mean? Can the above wisdom be turned into a theorem?

It is our purpose here to clarify these issues in a formal framework. Specifically, we seek *epistemic* conditions for a given strategy profile to be a Nash equilibrium: conditions involving what the players know or believe about one another–in particular, about payoff functions, strategy choices, decision procedures, and beliefs about these matters.[5]

Surprisingly, we will find that common knowledge of the payoff functions and of rationality are *never* needed; much weaker epistemic conditions suffice. Common knowledge of the strategies actually being played is also irrelevant. What *does* turn out to be relevant is common knowledge of the beliefs that the players hold about the strategies of the others; but here, too, common knowledge is relevant only when there are at least three players.

The main results are informally described in Section 2. The background for a formal presentation is given in Section 3; the underlying tool is that of an *interactive belief system*, which provides a framework for formulating epistemic conditions. Illustrations of such systems are given in Section 4. The formal statements and proofs of the main results, in Section 6, are preceded by some lemmas in Section 5. Sections 7 and 8 contain a series of counterexamples showing the results to be sharp; the examples–particularly those of Section 7–provide insight into the role played by the various epistemic conditions. Section 9 shows that the results apply to infinite as well as finite belief systems. Section 10 is devoted to a discussion of conceptual aspects and of the related literature. An appendix treats extensions and converses of the main results.

The reader wishing to understand just the main ideas should read Sections 2 and 7, and skim Sections 3 and 4.

2 Description of the Results

An event is called *mutual knowledge* if all players simply know it (to be distinguished from common knowledge, which also requires higher knowledge levels–knowledge about knowledge, and so on). Our first and simplest result is Theorem 6.1: *Suppose that each player is rational and knows his*

[2] An event is called *common knowledge* if all players know it, all know that all know it, and so on ad infinitum (Lewis [11, 1969]).

[3] See Section 10d for a discussion of why the payoff functions can be identified with the "structure of the game."

[4] I.e., that the players are optimizers; that given the opportunity, they will choose a higher payoff. A formal definition is given below.

[5] Other epistemic conditions for Nash equilibrium have been obtained by Armbruster and Boege [1, 1979] and Tan and Werlang [19, 1988].

own payoff function, and that the strategy choices of the players are mutually known. Then these choices constitute a Nash equilibrium in the game being played.

The proof is immediate: Since each player knows the choices of the others, and is rational, his choice must be optimal given theirs; so by definition, we are at a Nash equilibrium.

Note that neither the players' rationality, nor their payoff functions, nor their strategy choices are assumed common knowledge. For strategies, only mutual knowledge is assumed. For rationality and the structure of the game, not even mutual knowledge is assumed; only that the players are in fact rational, and know their own payoff functions.[6]

Theorem 6.1 applies to all pure strategy profiles. It applies also to mixed strategy profiles, under the traditional view of mixed strategies as conscious randomizations; in that case, of course, it is the mixture that must be mutually known, not just their pure realizations.

In recent years, a different view of mixed strategies has emerged.[7] In this view, players do not randomize; each player chooses some definite pure strategy. But the other players need not know which one, and the mixture represents their uncertainty, their probability assessment of his choice. This is the view adopted in the sequel; it fits in well with Bayesian approach to game theory, in which uncertainty about strategic choices of others is, like any other uncertainty, subject to probability assessment by each player.

For brevity, let us refer to pure strategies as *actions*. Define the *conjecture* of a player as his probability assessment of the actions of the other players. Call a player *rational* if his action maximizes his expected payoff given his conjecture.

When there are two players (and only then), the conjecture of each player is a mixed strategy of the other player. Because of this, the result in the two-person and n-person cases are quite different. For *two*-person games, we have the following (Theorem 6.2): *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.*

Theorem 6.2 differs from Theorem 6.1 in two ways. First, in both the conclusion and the hypothesis, strategy choices are replaced by conjectures; thus we get "conjectural equilibrium"–an equilibrium in conjectures, not in strategies actually played. Second, the hypothesis calls not just for the fact of rationality, but for mutual knowledge of this fact, and for mutual knowledge of the payoff functions. But common knowledge still does not enter the picture.

Since we are now viewing mixed strategies as conjectures, it is natural that conjectures replace choices in the result. So with n players, too, one might expect a theorem roughly analogous to Theorem 6.2; i.e., that mutual knowledge of the conjectures (when combined with appropriate assumptions about rationality and the payoff functions) is sufficient for them to be in equilibrium. But here we are in for a surprise: when $n > 2$, the conditions for a conjectural equilibrium become

[6] When a game is presented in strategic form, as here, knowledge of one's own payoff function may be considered tautologous. See Section 3.

[7] Harsanyi [8, 1973], Armbruster and Boege [1, 1979], Aumann [3, 1987], Tan and Werlang [19, 1988], Brandenburger and Dekel [5, 1989], among others.

much more stringent.

To understand the situation, note that the conjecture of each player i is a probability mixture of $(n-1)$-tuples of pure strategies of the other players. So when $n > 2$, it is not itself a mixed strategy; however, it induces a mixed strategy[8] for each player j other than i, called i's *conjecture about* j. One difficulty is that different players other than j may have different conjectures about j, in which case it is not clear how to define j's component of the conjectural equilibrium we seek to construct.

To present Theorem 6.3, the n-person "conjecture theorem", one more concept is needed. We say that the players have a *common prior*[9] if all differences between their probability assessments are due only to differences in their information; more precisely, if there is an outside observer O with no private information,[10] such that for all players i, if O were given i's information, his probability assessments would be the same as i's.

Theorem 6.3 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 players i agree on the same conjecture σ_j about j, and the resulting profile $(\sigma_1, \ldots, \sigma_n)$ of mixed strategies is a Nash equilibrium.*

The above three theorems give sufficient epistemic conditions for Nash equilibrium. The conditions are 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, all three theorems are "sharp," in the sense that they cannot be improved upon; none of the conditions can be dispensed with, or, so far as we can see, significantly weakened.

The presentation in this section, while correct, has been informal. For a formal presentation, one needs a framework for describing "epistemic" situations 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. Such frameworks are available in the differential information literature; a particular adaptation is presented in the next section.

3 Interactive Belief Systems

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

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

and for each type s_i of i,

[8] The marginal on j's strategy space of i's overall conjecture.

[9] Aumann [3, 1987]; for a formal definition, see Section 3. Harsanyi [7, 1967-68] uses term "consistency" to describe this situation.

[10] That is, what O knows is common knowledge among the players; each player knows everything that O knows.

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

(3.3) an action a_i for i (s_i's *action*),

and

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

The action sets A_i are assumed finite. One may also think of the spaces S_i as finite; the ideas are then more transparent. For a general definition, where the S_i are measurable spaces and the theories are probability measures,[11] see Section 9.

A belief system is a formal description of the players' beliefs–about each others' 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. See Section 10a for further discussion.

Set $S := S_1 \times \cdots \times S_n$. Call the members $s = (s_1, \ldots, s_n)$ of S *states of the world*, or simply *states*. An *event* is a subset E of S. Denote by $p(\cdot; s_i)$ the probability distribution on S induced by s_i's theory; formally, if E is an event, then $p(E; s_i)$ is the probability assigned by s_i to $\{s^{-i} \in S^{-i} : (s_i, s^{-i}) \in E\}$.

A function $g : A \to \mathbb{R}^n$ (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 $a^{-i} := (a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n)$. When referring to player i, the phrase "at s" means "at s_i". Thus, "i's action at s" means s_i's action (see (3.3)); we denote it $\mathbf{a}_i(s)$, and write $\mathbf{a}(s)$ for the n-tuple $(\mathbf{a}_1(s), \ldots, \mathbf{a}_n(s))$ of actions at s. Similarly, "i's payoff function at s" means s_i's payoff function (see (3.4)); we denote it $\mathbf{g}_i(s)$, and write $\mathbf{g}(s)$ for the n-tuple $(\mathbf{g}_1(s), \ldots, \mathbf{g}_n(s))$ of payoff functions[12] 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 $\mathbf{a}_i(s)$, $\mathbf{a}(s)$, $\mathbf{g}_i(s)$, and $\mathbf{g}(s)$) 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 $\{s \in S : \mathbf{x}(s) = x\}$. For example, $[a_i]$ denotes the event that i chooses the action a_i; and $[g]$ denotes the event that the game g is being played.

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 $\boldsymbol{\varphi}^i(s)$, called *$i$'s conjecture at s*, given by $\boldsymbol{\varphi}^i(s)\left(a^{-i}\right) := p\left(\left[a^{-i}\right]; s_i\right)$. We denote the n-tuple $\left(\boldsymbol{\varphi}^1(s), \ldots, \boldsymbol{\varphi}^n(s)\right)$ of conjectures at s by $\boldsymbol{\varphi}(s)$.

[11] Readers unfamiliar with measure theory may think of the type spaces S_i as finite throughout the paper. All the examples involve finite S_i only. The results, too, are stated and proved without reference to measure theory, and may be understood completely in terms of finite S_i. On the other hand, we do not *require* finite S_i; the definitions, theorems, and proofs are all worded so that when interpreted as in Section 9, they apply without change to the general case. One can also dispense with finiteness of the action spaces A_i; but that is both more involved and less important, and we will not do it here.

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

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 $h_i := \mathbf{g}_i(s)$ and $b_i := \mathbf{a}_i(s)$, this means that $\text{Exp}\left(h_i\left(b_i, \mathbf{a}^{-i}\right)|s_i\right) \geq \text{Exp}\left(h_i\left(a_i, \mathbf{a}^{-i}\right)|s_i\right)$ for all a^i in A^i. Another way of saying this is that i's actual choice b_i maximizes the expectation of his actual payoff h_i when the other players' actions are distributed according to his actual conjecture $\boldsymbol{\varphi}^i(s)$.

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 $K^1E := K_1E \cap \cdots \cap K_nE$; thus K^1E is the event that all player know E. If $s \in K^1E$, call E *mutually known at* s. Set $CKE := K^1E \cap K^1K^1E \cap K^1K^1K^1E \cap \cdots$; 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,

(3.5) 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, (3.5) says that for each player i, if two events have proportional probabilities given any s_i, then they have proportional prior probabilities.

Another regularity condition that is sometimes used in the differential information literature is "mutual absolute continuity." We do not define this here because we have no use for it.

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. Thus at s, Rowena may know E, but not know that Colin knows E; or at s, it may be that Colin is rational, and that Rowena knows this, but that Colin does not know that Rowena knows it. Some illustrations of these ideas are given in the next section.

4 Illustrations

Example 4.1 *We start with a belief system in which all types of each player i have the same payoff function g_i, namely that depicted in Figure 4.1. Thus the game being played is commonly known. Call the row and column players (Players 1 and 2) "Rowena" and "Colin" respectively.*

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

Figure 4.1

The theories are depicted in Figure 4.2; 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 $1/2 - 1/2$ probabilities to Rowena's type being D_1 or D_2. So at 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.

	c_1	d_1	d_2
C_1	2/5, 2/3	3/5, 3/5	0, 0
D_1	1/2, 1/3	0, 0	1/2, 1/2
D_2	0, 0	2/3, 2/5	1/3, 1/2

Figure 4.2

We have here a typical instance of Theorem 6.1 (see the beginning of Section 2), which also shows that the folk wisdom cited in the introduction is misleading. 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. Thus, though Colin knows that Rowena will play D, she doesn't know that he knows this; indeed, she attributes probability $2/3$ to his attributing probability $3/5$ 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 $2/5$, rather than c, with an expected payoff of $6/5$; thus this type is "irrational." At (D_2, d_2), Rowena attributes probability $2/3$ to Colin being of this irrational type.

	c_1	d_1	d_2
C_1	0.2	0.3	0
D_1	0.1	0	0.1
D_2	0	0.2	0.1

Figure 4.3

Note that the players have a common prior (Figure 4.3). But Theorem 6.1 has nothing to do with common priors. If, for example, we change the theory of Rowena's type D_2 from $\frac{2}{3}d_1 + \frac{1}{3}d_2$ to $\frac{1}{2}d_1 + \frac{1}{2}d_2$, then there is no longer a common prior; but (D, d) is still Nash equilibrium, for the same reasons as above. (As usual, $\frac{2}{3}d_1 + \frac{1}{3}d_2$ stands for the $\frac{2}{3} - \frac{1}{3}$ probability combination of d_1 and d_2. Similar notation will be used throughout the sequel.)

Example 4.2 *As in the previous example, the game being played here–"matching pennies" (Figure 4.4)–is commonly known. The theories are depicted in Figure 4.5. At the state (H_1, h_1), the conjectures of Rowena and Colin are $\frac{1}{2}h + \frac{1}{2}t$ and $\frac{1}{2}H + \frac{1}{2}T$ respectively, and these conjectures are mutually known (i.e., each knows that these are the conjectures). Moreover, the rationality of both*

players is mutually known. Thus Theorem 6.2 (Section 2) implies that $\left(\frac{1}{2}H + \frac{1}{2}T, \frac{1}{2}h + \frac{1}{2}t\right)$ *is a Nash equilibrium, which indeed it is.*

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

Figure 4.4

	h_1	t_1	t_2
H_1	1/2, 1/2	1/2, 1/2	0, 0
T_1	1/2, 1/2	0, 0	1/2, 1
T_2	0, 0	1, 1/2	0, 0

Figure 4.5

Note that neither the conjectures of the players nor their rationality are commonly known. Indeed, at (T_1, t_2), Colin knows that Rowena plays T, so that his conjecture is not $\frac{1}{2}H + \frac{1}{2}T$ but T; so it is irrational for him to play t, which yields him 0, rather than 1 that he could get by playing h. At the state (H_1, h_1), Colin attributes probability 1/2 to Rowena attributing probability 1/2 to the state (T_1, t_2); so at (H_1, h_1), there is common knowledge neither of Colin's conjecture $\frac{1}{2}H + \frac{1}{2}T$, nor of his rationality.

	h_1	t_1	t_2
H_1	0.2	0.2	0
T_1	0.2	0	0.2
T_2	0	0.2	0

Figure 4.6

Note, too, that like in the previous example, this belief system has a common prior (Figure 4.6). But also like there, this is not essential; the discussion would not be affected if, say, we changed the theory of Rowena's type T_2 from t_1 to $\frac{1}{2}t_1 + \frac{1}{2}t_2$.

5 Properties of Belief Systems

In this section we formally establish some basic properties of belief systems, which are needed in the sequel. These properties are intuitively fairly obvious, and some are well-known in various formalizations of interactive knowledge theory; so this section can be omitted at a first reading.

Lemma 5.1 *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, $F := \{t \in S : p(E; t_i) = \pi\}$. If $s \in F$, then $p(E; s_i) = \pi$, so all states u with $u_i = s_i$ are in F. Therefore $p(F; s_i) = 1$; that is, i knows F at s, so $s \in K_i F$.

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. ■

Corollary 5.1 *Let φ be n-tuple of conjectures. Suppose that at some state s, it is mutually known that $\varphi = \varphi$. Then $\varphi(s) = \varphi$. (In words: if it mutually known that the conjectures are φ, then they are indeed φ.)*

Corollary 5.2 *A player is rational if and only if he knows that he is rational.*

Corollary 5.3 *$K_i K_i E = K_i E$ (a player knows something if and only if he knows that he knows it), and $K_i \neg K_i E = \neg K_i E$ (a player doesn't know something if and only if he knows that he doesn't know it).*

Lemma 5.2 *$K_i(E_1 \cap E_2 \cap \cdots) = K_i E_1 \cap K_i E_2 \cap \cdots$ (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 $E_1 \cap E_2 \cap \cdots$ if and only if he ascribes probability 1 to each $E_1, E_2, \ldots$. ■

Lemma 5.3 *$CKE \subset K_i CKE$ (if something is commonly known, then each player knows that it is commonly known).*

Proof. Since $K_i K^1 F \supset K^1 K^1 F$ for all F, Lemma 5.2 yields $K_i CKE = K_i\left(K^1 E \cap K^1 K^1 E \cap \cdots\right) = K_i K^1 E \cap K_i K^1 K^1 E \cap \cdots \supset K^1 K^1 E \cap K^1 K^1 K^1 E \cap \cdots \supset CKE$. ■

Lemma 5.4 *Suppose P is a common prior, $K_i H \supset H$, 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 $K_i H \supset H$ it follows that $p(H; s_i) = 1$ or 0 according as to whether s_i is or is not[13] 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 (3.5). ■

The following lemma is not needed for the proofs of the theorems, but relates to the examples in Section 7. Set $K^2 E := K^1 K^1 E$, $K^3 E := K^1 K^2 E$, and so on. If $s \in K^m E$, call E *mutually known of order m* at s.

Lemma 5.5 *$K^m E \subset K^{m-1} E$ for $m > 1$ (mutual knowledge of order m implies mutual knowledge of orders $1, ..., m-1$).*

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

Proof. By Lemma 5.2 and Corollary 5.3, $K^2E = K^1K^1E = \cap_{i=1}^n K_iK^1E \subset K_iK^1E = K_i \cap_{j=1}^n K_jE \subset K_iK_iE$. Since this is so for all i, we get $K^2E \subset \cap_{i=1}^n K_iE = K^1E$. The result for $m > 2$ follows from this by substituting $K^{m-2}E$ for E. ■

6 Formal Statements and Proofs of the Theorems

We now formally state and prove Theorems 6.1-6.3. For more transparent paraphrases (using the same terminology), see Section 2.

Theorem 6.1 *Let a be n-tuple of actions. Suppose that at some state s, all players are rational, and all know that $\mathbf{a} = a$. Then a is a Nash equilibrium.*

Proof. Immediate (see beginning of Section 2). ■

Theorem 6.2 *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 $\boldsymbol{\varphi} = \varphi$. Then $\left(\varphi^2, \varphi^1\right)$ is Nash equilibrium of g.*

The proof uses a lemma; we state it for the n-person case, since it is needed again in the proof of Theorem 6.3.

Lemma 6.1 *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 $\boldsymbol{\varphi} = \varphi$. 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*[14] *φ^j.*

Proof. By Corollary 5.1, 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$]$, $\left[\varphi^j\right]$, and $[g_j]$. When one of four events has positive probability, and the other three each have probability 1, then their intersection is non-empty. 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. ■

Proof of Theorem 6.2. By Lemma 6.1, 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 $\left(\varphi^2, \varphi^1\right)$ is a Nash equilibrium of g. ■

Theorem 6.3 *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 $\boldsymbol{\varphi} = \varphi$. Then for each j, all the conjectures φ^i of players i other than j induce the same conjecture σ_j for j, and $(\sigma_1, \ldots, \sigma_n)$ is a Nash equilibrium of g.*

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

The proof requires a lemma.

Lemma 6.2 *Let Q be a probability distribution on A with*[15] $Q(a) = Q(a_i)\,Q\left(a^{-i}\right)$ *for all a in A and all i. Then $Q(a) = Q(a_1)\cdots Q(a_n)$ for all a.*

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

Proof of Theorem 6.3. Set $F := CK[\varphi]$, 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,

$$Q(a) = Q(a_i)\,Q\left(a^{-i}\right). \tag{6.1}$$

Set $H := [a_i] \cap F$. By Lemmas 5.2 and 5.3, $K_i H \supset H$, since i knows his own action. If $s \in H$, it is commonly, and so mutually known at s that $\boldsymbol{\varphi} = \varphi$; so by Corollary 5.1, $\boldsymbol{\varphi}(s) = \varphi$; that is, $p\left(\left[a^{-i}\right];s_i\right) = \varphi^i\left(a^{-i}\right)$. So Lemma 5.4 (with $E = \left[a^{-i}\right]$) yields $P([a]\,|F) = P\left(\left[a^{-i}\right]\cap H\right) = \varphi^i\left(a^{-i}\right)P(H) = \varphi^i\left(a^{-i}\right)P([a_i]\,|F)$. Dividing by $P(F)$ yields $Q(a) = \varphi^i\left(a^{-i}\right)Q(a_i)$; then summing over a_i, we get

$$Q\left(a^{-i}\right) = \varphi_i\left(a^{-i}\right). \tag{6.2}$$

Thus $Q(a) = Q\left(a^{-i}\right)Q(a_i)$, which is (6.1).

For each j, define a probability distribution σ_j on A_j by $\sigma_j(a_j) := Q(a_j)$. Then (6.2) yields $\varphi^i(a_j) = Q(a_j) = \sigma_j(a_j)$ for $j \neq i$. Thus for all i, the conjecture for j induced byφ^i is σ_j, which does not depend on i. Lemma 6.2, (6.1), and (6.2) then yield

$$\varphi^i\left(a^{-i}\right) = \sigma_1(a_1)\cdots\sigma_{i-1}(a_{i-1})\,\sigma_{i+1}(a_{i+1})\cdots\sigma_n(a_n); \tag{6.3}$$

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

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 players are rational, and that $\boldsymbol{\varphi} = \varphi$. So by Lemma 6.1, each action a_j with $\varphi^i(a_j) > 0$ for some $i \neq j$ maximizes g_j against φ^j. By (6.3), these a_j are precisely the ones that appear with positive probability in σ_j. Again using (6.3), 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. ■

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

7 The Main Counterexamples

This section explores possible variations on Theorem 6.3 (the result giving sufficient epistemic conditions, when $n \geq 3$, for the players' conjectures to yield a Nash equilibrium). For simplicity, let $n = 3$. 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 in 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 7.1 shows that this hope is vain, even when the priors are common and rationality is 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 of a high order?

Again, the answer is no. In Example 7.2, there is mutual knowledge of overall conjectures (which may be of arbitrarily high order), common knowledge of rationality, and common prior, but the individual conjectures do not constitute a Nash equilibrium.

What drives this example is that different players have different individual conjectures about some particular player j, so there isn't even even a clear *candidate* for a Nash equilibrium.[16] This raises the question of whether (sufficiently high order) mutual knowledge of the overall conjectures implies Nash equilibrium of the individual conjectures when the players do happen to agree, in *addition* to assuming (sufficiently high order) mutual knowledge of the overall conjectures. Do we get Nash equilibrium?

Again, the answer is no; this is shown in Example 7.3.

Finally, Example 7.4 shows that the common prior assumption is really needed; it exhibits a situation with common knowledge of the overall conjectures and of rationality, where 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 common priors, will do.

Except in Example 7.4, 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 (like in Section 4) 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 4, each

[16] It is *not* what drives Example 7.1; since the individual conjectures are commonly known there, they must agree (Aumann [2, 1976]).

type is denoted by the same letter as its action, and a subscript is added.

Example 7.1 *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. Consider the game of Figure 7.1, with theories induced by the common prior in Figure 7.2.*

	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 7.1

	L_1	R_1
U_1	1/4	0
D_1	0	1/4

W_1

	L_1	R_1
U	0	1/4
D	1/4	0

E_1

Figure 7.2

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 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* $\left(\frac{1}{2}U + \frac{1}{2}D, \frac{1}{2}L + \frac{1}{2}R, \frac{1}{2}W + \frac{1}{2}E\right)$ *is not a Nash equilibrium, because if these were independent mixed strategies, Rowena could gain by moving to* D*.*

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* $\left(\frac{1}{2}LW + \frac{1}{2}RE\right)$*, but nobody else knows that that is her conjecture.*

Example 7.2 *Here we have mutual knowledge of the overall conjectures, common knowledge of rationality, and common priors; yet individual conjectures don't agree, so one can't even identify a candidate for a Nash equilibrium.*

	L_1	L_2	L_3
U_1	0.4 W_1	0.2 E_1	
U_2		0.2 W_2	0.1 E_2
U_3			0.1 W_3

Figure 7.3

Consider a three-person game in which Rowena and Colin each have just one action–say U *for*

Rowena and L for Colin–and Matt has two actions, W and E. The payoffs are unimportant in this example,[17] *since we are only interested in showing that the individual conjectures do not agree. Let the belief system be as in Figure 7.3. As usual, Rowena's types are depicted by rows, Colin's by columns. Matt's types are indicated in the individual boxes in the diagram; note that in this case, he knows the true state.*

Consider the state (U_2, L_2, W_2)*, or simply* W_2 *for short. At this state, Rowena's conjecture is* $\frac{2}{3}LW + \frac{1}{3}LE$*, Colin's is* $\frac{1}{2}UW + \frac{1}{2}UE$*, and Matt's is* UL*. Rowena knows Colin's and Matt's conjectures, as they are the same at the only other state* (E_2) *that she considers possible.*[18] *Similarly, Colin knows Rowena's and Matt's conjectures, as they are the same at the only other state* (E_1) *that he considers possible. Matt knows Rowena's and Colin's conjectures, since he knows that the true state is* W_2*. So the conjectures are mutually known. Yet Rowena's conjecture for Matt,* $\frac{2}{3}W + \frac{1}{3}E$*, is different from Colin's,* $\frac{1}{2}W + \frac{1}{2}E$*.*

The same idea yields examples of this kind with higher order mutual knowledge of the conjectures. See Figure 7.4. At W_3, there is third order mutual knowledge of the conjectures. By lengthening the staircase[19] and choosing a state in its middle, one can get mutual knowledge of arbitrarily high order that Rowena's conjecture for Matt is $\frac{2}{3}W + \frac{1}{3}E$, while Colin's is $\frac{1}{2}W + \frac{1}{2}E$.

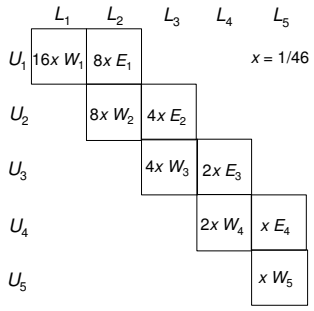

Figure 7.4

The perspicacious reader will have realized that this example is not intrinsically game-theoretic. It really boils down to a question about "agreeing to disagree:" Suppose that two people with the same prior get different information, and that given this information, their posterior probabilities for some event A are mutual knowledge of some order. Are the posterior probabilities then equal? (Here the individuals are Rowena and Colin, and the event A is that Matt chooses W; but actually that's just window dressing.) An example in Aumann [2, 1976] provides a negative answer to the question in the case of first order mutual knowledge. Geanakoplos and Polemarchakis [6, 1982] showed that the answer is negative also for higher order mutual knowledge. The ingenious example

[17] They can easily be chosen so that rationality is common knowledge.

[18] I.e., to which she assigns positive probability.

[19] Alternatively, one can use a single probability system with countably many states, represented by a staircase anchored at the top left and extending infinitely downwards to the right (Jacob's ladder?). By choosing a state sufficiently far from the top, one can get mutual knowledge of any given order.

presented here is due to John Geanakoplos.[20]

Example 7.3 *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 7.5. For Rowena and Colin, this is simply "matching pennies" (Figure 4.4); 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* $3/2$ *for* W*, and* 2 *for* E*; so he must play* E*. Hence* $\left(\frac{1}{2}H + \frac{1}{2}T, \frac{1}{2}h + \frac{1}{2}t, E\right)$ *is the unique Nash equilibrium of this game.*

	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 7.5

Consider now the theories induced by the common prior in Figure 7.6. Rowena and Colin know which of the three "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.*

	h_1	t_1	h_2	t_2	h_3	t_3
H_1	$9x\ W_1$	$9x\ E_1$				$x = 1/52$
T_1	$9x\ E_1$	$9x\ W_1$				
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 7.6

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 7.6; 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, the probability is three times that on the right. So Matt assigns*

[20] Private communication. The essential difference between the 1982 example of Geanakoplos and Polemarchakis and the above example of Geanakoplos is that in the former, Rowena's and Colin's probabilities for A approach each other as the other m of mutual knowledge approaches ∞, whereas in the latter, they remain at $2/3$ and $1/2$ no matter how large m is.

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* $\frac{3}{8}3 + \frac{3}{8}3 + \frac{1}{8}0 + \frac{1}{8}0 = 2\frac{1}{4}$, *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 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 for each other. Also, 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.*

To understand the construction, note that in the east matrix, Matt gets 2 no matter what the other do; in the west matrix, he gets 3 if the others play on the (main) diagonal, 0 if they play off. Thus he is motivated to play W when his conjecture assigns a high probability to their playing on the diagonal. The common prior in Figure 7.6 generates such a conjecture that is mutually known, in spite of the overall probability of the diagonal being only 1/2. The reason for using the diagonal[21] is to avoid correlation with Rowena's or Colin's information, so as to assure that Matt's conjectures about them will agree with their conjectures about each other.

As in Example 7.2, one can construct similar example for which the mutual knowledge of the conjectures is of arbitrarily high order, simply by using more boxes; the result follows as before.

Agreement between individual conjectures may be viewed as a kind of "consistency" between the overall conjectures φ^i. A stronger consistency condition,[22] suggested by Kenneth Arrow, is that there be a single probability distribution ϕ on $A = \times_{i=1}^n A_i$ whose marginals on the spaces A^{-i} are the φ^i. One may ask whether with this stronger condition, one can replace common by mutual knowledge in Theorem 6.3. The answer is still "no", even with mutual knowledge of arbitrarily high order. The above example satisfies Arrow's condition, with $\phi = \frac{3}{8}HhW + \frac{3}{8}TtW + \frac{1}{8}ThW + \frac{1}{8}HtW$

[21] Rather than the upper row, say.

[22] Though similar in form, this condition neither implies nor is implied by common priors. We saw in Example 7.2 that common priors do not even imply agreement between individual forecasts; a fortiori, they do not imply Arrow's condition. In the opposite direction, Example 7.4 satisfies Arrow's condition, but has no common prior.

(Figure 7.7).

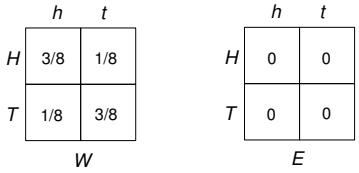

Figure 7.7

Example 7.4 *Here we show that one cannot dispense with common priors in Theorem 6.3. Consider again the game of Figure 7.5, with the theories depicted in Figure 7.8 (presented in the style of Figure 4.2 and 4.5; note that Matt has no type*[23] *whose action is E). At each state there is common knowledge of rationality, of overall conjectures (which are the same as in the previous example), and of the game. As before, Arrow's condition is satisfied, and it follows that the individual conjectures are in agreement. And as before, the individual conjectures* $\left(\frac{1}{2}H + \frac{1}{2}T, \frac{1}{2}h + \frac{1}{2}t\right)$ *do not constitute a Nash equilibrium.*

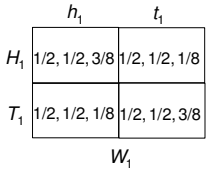

Figure 7.8

8 Additional Counterexamples

In the previous section we saw that the assumption of a common prior and of common knowledge of the conjectures are essential in Theorem 6.3. This section explores the assumption of mutual knowledge of rationality and of the game being played (in both Theorems 6.2 and 6.3), and shows that they, too, cannot be substantially weakened.

Example 8.1 *Here we show that in Theorems 6.2 and 6.3, mutual knowledge of rationality cannot be replaced by the simple fact of rationality (as in Theorem 6.1). Consider again the game of "matching pennies" (Figure 4.4), this time with theories induced be the common prior in Figure 8.1.*

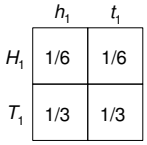

Figure 8.1

At the state (H_1, h_1)*, Colin's and Rowena's conjectures are commonly known to be* $\frac{1}{3}H + \frac{2}{3}T$

[23] If one wishes, one can introduce a type E_1 of Matt to which Rowena's and Colin's types ascribe probability 0, and whose theory is, say, $\frac{1}{4}Hh + \frac{1}{4}Tt + \frac{1}{4}Th + \frac{1}{4}Ht$.

and $\frac{1}{2}h + \frac{1}{2}t$ *respectively, and both are in fact rational (indeed Rowena's rationality is commonly known); but* $\left(\frac{1}{3}H + \frac{2}{3}T, \frac{1}{2}h + \frac{1}{2}t\right)$ *is not a Nash equilibrium, since the only equilibrium of "matching pennies" is* $\left(\frac{1}{2}H + \frac{1}{2}T, \frac{1}{2}h + \frac{1}{2}t\right)$*. Note that at the state* (H_1, h_1)*, Rowena does not know that Colin is rational.*

Example 8.2 *Here we show that in Theorems 6.2 and 6.3, knowing one's own payoff function does not suffice; one needs mutual knowledge of all the payoff functions. Consider a two-person belief system where one of two games,* T *(top) or* B *(bottom), is being played (Figure 8.2).*

Game T

	S
U	1, 1
D	0, 0

Game B

	S
U	0, 1
D	1, 0

Figure 8.2

	S_1
TU_1	1/2
BD_1	1/2

Figure 8.3

The theories are given by the common prior in Figure 8.3. Thus Rowena knows Colin's payoff function, but Colin does not know Rowena's. Rowena's type TU_1 *has the Top payoff function and plays Up, whereas* BD_1 *has the Bottom payoff function and plays Down. Colin has just a single type,* S_1*. At both states, both players are rational: Rowena, who knows the game, always plays an action that is strictly dominant in the true game; Colin has no choice, so what he does is rational. So there is common knowledge of rationality. At both states, Colin's conjecture about Rowena is* $\frac{1}{2}U + \frac{1}{2}D$*; Rowena's conjecture about Colin is* S*. But* $\left(\frac{1}{2}U + \frac{1}{2}D, S\right)$ *is not a Nash equilibrium in either of the games; Rowena prefers* U *in the top game,* D *in the bottom game.*

Example 8.3 *Here we show that the hypotheses of Theorem 6.3 do not imply that rationality is commonly known (as is the case when the game* g *is commonly known–see Proposition A.1). Consider a two-person belief system where one of the two games of Figure 8.4 is being played. The theories are given by the common prior in Figure 8.5. Of Rowena's three types,* TU_1 *and* TD_1 *are rational, whereas* BD_1 *(who plays Down in Game B) is irrational. Colin's two types,* S_1 *and* S_2*, differ only in their theories, and both are rational. The conjectures* $\frac{1}{2}U + \frac{1}{2}D$ *and* S *are common knowledge at all states. At the state* (TU_1, S_1)*, it is mutually known that* T *is the game being played,*

and that both players are rational; but neither rationality nor the game being played are commonly known.

Game T

	S
U	0, 1
D	0, 0

Game B

	S
U	1, 1
D	0, 0

Figure 8.4

	S_1	S_2
TU_1	1/4	1/4
TD_1	1/4	0
BD_1	0	1/4

Figure 8.5

9 General (Infinite) Belief Systems

For a general definition of a belief system, we specify that the type spaces S_1 be measurable spaces. As before, a *theory* is a probability measure on $S^{-i} = \times_{j \neq i} S_j$, which is now endowed with the standard product structure.[24] 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" $\mathbf{a}_i$ ((3.3)) are assumed measurable; so are the payoff functions $\mathbf{g}_i$ ((3.4)), as functions of s_i, for each action n-tuple a separately. Also the "theory functions" ((3.2)) are assumed measurable, in the sense that for each event E and player i, the probability $p(E; s_1)$ is measurable as a function of the type s_i. It follows that also the conjectures $\boldsymbol{\varphi}^i$ are measurable functions of s_i.

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

10 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

[24] 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.

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 [16, 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. 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, every-day 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 prescribing what a player must do at a state takes away his freedom. This is nonsensical; the player may do what he wants. It is simply that whatever he does is part of the description of the state. If he wishes to do something else, he is heartily welcome to do it, but he thereby changes the state.

Historically, belief systems were introduced by John Harsanyi [7, 1967-68], to enable a coherent formulation of games in which the players need not know each others' payoff functions. To analyze such games, it is not enough to specify each player's beliefs about (i.e., probability distributions on) the payoff functions of the others; one must also specify the beliefs of the players about the beliefs of the players about the payoff functions, the beliefs of the players about *these* beliefs, and so on ad infinitum. This complicated infinite regress seemed to make useful analysis very difficult.

Harsanyi's ingenious solution was to think of each player as being one of several possible "types," where a type determines both a player's own payoff function and a belief about the types of the others. The belief of a player about the types of the others induces a belief about their payoff functions; it also induces a belief about their beliefs about the types, and so a belief about the beliefs about the payoff functions. The reasoning continues indefinitely. Thus from an *I-game* ("I" for incomplete information), as Harsanyi calls his type-model, one can read off the entire infinite regress of beliefs.

Belief systems as defined in Section 3 are formally just like Harsanyi's I-games, except that in belief systems, a player's type determines his action as well as his payoff function and his belief about other players' types.[25] As above, it follows that the player's type determines his entire *belief hierarchy*–i.e., the entire infinite regress of his beliefs about actions, beliefs about beliefs about actions, and so on–in addition to the infinite regress of beliefs about payoff functions, and how these two kinds of beliefs affect each other.

Traditionally, payoff functions have been treated as exogenous, actions as endogenous. It was thought that unlike payoff functions, actions should be "predicted" by the theory. Belief systems wipe out this distinction; they treat uncertainty about actions just like uncertainty about payoff

[25] For related ideas, see Armbuster and Boege [1, 1979], Boege and Eisele [4, 1979], and Tan and Werlang [19, 1988].

functions. Indeed, in this paper the focus is on actions;[26] uncertainty about payoff functions was included as an afterthought, because we realized that more comprehensive results can be obtained at almost no cost in the complexity of the proofs.

b. Common Knowledge of the Belief System Is the belief system itself common knowledge among players? If so, how does it get to be common knowledge? If not, how do we take into account the players' uncertainty about it?

A related question is whether the belief system is exogenous, like a game or a market model. If not, where does it come from?

The key to these issues was provided[27] in a fundamental paper of Mertens and Zamir [13, 1985]. They treat the Harsanyi case, in which the "underlying variables"[28] are the payoff functions only (see (a) above); but the result apply without change to our situation, where actions as well as payoff functions are underlying variables.

At (a) above, we explained how each type in a belief system determines a belief hierarchy. Mertens and Zamir reverse this procedure: They start with the belief hierarchies, and construct belief systems from them. Specifically, they define the *universal belief space* as a belief system in which the type space of each player is simply the set of *all* his belief hierarchies that satisfy certain minimal consistency conditions. Thus the universal belief space is not exogenously given, like a game or market; rather, it is an analytic tool, like payoff matrix of a game originally given in extensive form. It follows that the universal belief space may be considered common knowledge.

Though the universal belief space is infinite, it is the disjoint union of infinitely many "common knowledge components,"[29] many of which are finite. Mertens and Zamir call any union of such components a *subspace* of the universal belief space. It follows that when the belief system is a subspace, then *the belief system itself is common knowledge.*

It may be shown that *any* belief system $\mathcal{B}$ for $A_1 \times \cdots \times A_n$–including, of course, a finite one–is "isomorphic" to a subspace of the universal belief space.[30] From all this we conclude that the belief system itself may *always* be considered common knowledge.

c. 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. In the formalism of belief systems[31] (Section 3), absolute certainty has little concrete meaning; a player can be absolutely certain only on his own action, his own payoff function, and his own theory, not of anything pertaining to anybody else.

Choosing between the terms "know" and "believe" caused us many sleepless nights. In the end,

[26] As is apparent from the examples in Sections 4, 7, and 8, in most of which the game g is common knowledge.

[27] See also Armbuster and Boege [1, 1979] and Boege and Eisele [4, 1979].

[28] The variables about which beliefs–of all orders–are held.

[29] At each state s of such a component S, the identity of that component is commonly known (i.e., it is commonly known at s that the true state is in S, though it need not be commonly known that the true state is s).

[30] This is because each type in $\mathcal{B}$ determines a belief hierarchy (see (a)). The set of n-tuples of all these belief hierarchies is the subspace of the universal belief space that is isomorphic to $\mathcal{B}$.

[31] Unlike that of partitions (see (e) below).

we decided on "know" because it enables simpler, less convoluted language, and because we were glad to de-emphasize the relatively minor conceptual difference between probability 1 and absolute certainty.[32]

Note that since our conditions are sufficient, our 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.

d. Structure of the Game In Section 1, we identified the "structure of the game" with n-tuple g of payoff functions. As pointed out by John Harsanyi [7, 1967-68], this involves no conceptual loss of generality, as one can always pick the action sets A_j sufficiently large to justify this.

e. Alternative Formalisms Instead of using belief systems to formulate and prove our results and examples, one may use knowledge partitions (e.g., Aumann [2, 1976]). The advantage of the partition formalism is that with it one can represent absolute certainty, not just probability 1 knowledge (see (c)). Also. some of the proofs may become marginally simpler.[33] On the other hand, the partition formalism–especially when combined with probabilities–is itself more complex than that of belief systems, and it is desirable to use as simple, transparent, and unified an approach as possible.

f. Independent Conjectures Theorem 6.3 deduces independence of the individual conjectures from common knowledge of the overall conjectures. An alternative approach is embodied in the following:

Remark 10.1 *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.*

The proof is as in the last part of Theorem 6.3's proof, after (6.3) has been established.

Here we assume mutual rather than common knowledge of conjectures and do not assume common priors. 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. Because of the strength of these assumptions, the result is of limited interest.

Mutual independence of the conjectures of a player i about different players j seems a particularly untenable assumption. It may well be that Colin and Matt "choose their actions independently" in the sense of not consulting each other or even communicating, but that by no means implies that Rowena's conjecture about Colin is stochastically independent of her conjecture about Matt. It is possible to have considerable stochastic dependence even when there is no possibility of communication or correlation between Colin and Matt–even when they are placed in separate rooms and then

[32] Another reason is that "belief" often refers to a general probability distribution, which does not go well with using "know" to mean "ascribe probability 1 to".

[33] This is natural, as the results are slightly weaker (see (c)).

informed of the game they are to play. The probability is in Rowena's head, it is a question of *her* beliefs, *her* ignorance or knowledge, it is not directly determined by some kind of physical process such as "independently" choosing or being in separate rooms or anything like that. Suppose, for example, that Colin and Matt could each act either "boldly" or "carefully." Even if Rowena knows "nothing at all" about Colin and Matt (whatever that might mean, perhaps she believes they came from unspecified different directions in outer space), her probability that Matt acts boldly given that Colin does might well be higher than her unconditional probability. After all, it is a question of *her* outlook on life, how *she* thinks people react to things; if she is at all sophisticated, she might easily say to herself, "well, I'm not sure of human nature, I don't know what people do in this kind of situation; but if Colin plays boldly, that tells me something about people in general, and it raises my probability that Matt will also play boldly." In fact, it is quite unlikely that the conjectures would be stochastically independent, one would have great difficulty in constructing a set of circumstances that would justify such a conclusion.

When teaching probability, we are at pains to point out that a sequence of coin tosses is *not* in general represented by an i.i.d. sequence of random variables. This would be the case only if the parameter of the coin were known with certainty before the first toss, a highly unlikely proposition. And this in spite of the fact that the tosses themselves are of course "physically" independent (whatever that may mean; we're not sure it means anything). The relation between "physical" and stochastic independence is murky at best.

Independence of individual conjectures is an appropriate assumption when one thinks of mixed strategies in the old way, in terms of explicit randomizations. With that kind of interpretation, "acting independently" is of course closely associated with stochastic independence. But not when the probabilities represent other players' ignorance.

To be sure, common knowledge of the conjectures is also a rather strong assumption. Moreover, we do *conclude* that the individual conjectures are independent. But there is a difference between an assumption that is merely strong and one that is groundless. More to the point, there is a qualitative difference between assuming common knowledge of the conjectures and assuming independence. Common knowledge of the conjectures describes what might almost be called a "physical" state of affairs. It might, for example, be the outgrowth of experience or learning, like common knowledge of a language. Independence, on the other hand, is in the mind of the decision maker. It isn't reasonable to make such an assumption when one can't describe some clear set of circumstances under which it would obtain–and that is very difficult to do. By contrast, common knowledge of the conjectures may itself be considered a "clear set of circumstances."

g. Epistemic "Equivalence" Conditions This paper is devoted to sufficient epistemic conditions for Nash equilibrium. Another kind of epistemic result is illustrated by the following:

Remark 10.2 *Let a be n-tuple of actions. Suppose that at some state s, it is commonly known that $\mathbf{g} = g$ and $\mathbf{a} = a$. Then rationality is commonly known at s iff a is a Nash equilibrium in g.*

For another result of this kind, which relates to Theorem 6.2 as the above relates to Theorem 6.1, see Brandenburger and Dekel (1989). These results provide sufficient conditions for the *equivalence* between Nash equilibrium and common knowledge of rationality, rather than directly for Nash equilibrium.

h. Global Epistemic Conditions The epistemic conditions discussed in this paper are *local*; they say when the players' actions or conjectures at a specific state constitute an equilibrium. In contrast, one can treat *global* conditions, which refer to the belief system as a whole. These are usually quite different from local conditions treated here. For example, the condition for correlated equilibrium in Aumann [3, 1987] is global.

A global epistemic condition for Nash equilibrium[34] is as follows:

Remark 10.3 *Suppose that the game g is fixed,*[35] *that there is a common prior P, and that all players are rational at all states. Then the distribution Q of action n-tuples is a Nash equilibrium in g if and only if for each player i, the expectation of i's conjecture given one of his actions a_i is the same for all a_i that are assigned positive probability by Q.*

Roughly speaking, the condition says that if the players use only information that is relevant to their choice of an action, then their conjectures are always the same, independent of their information.

i. Irrationality, Probability Assessments, and Payoff Functions There is a difficulty when explicitly allowing for irrationality, as we do here. The problem rears its ugly head on several levels. Most fundamental, a player is by definition rational if his choice maximizes his utility function. Again by definition, the utility function represents his preferences, and yet again by definition, the player prefers x to y iff he chooses x when given the choice between x and y. So "rationality" appears tautologous: utility is always maximized; it is *defined* as a function that the choice maximizes.

One may attempt to resolve this by noting that whereas rationality requires the choice x to maximize the utility function, that does not mean that any x maximizes some utility function. Indeed there may not be any utility function at all. Rationality must be understood in terms of a richer structure, one involving preferences (choices between pairs or from other subsets of the available alternatives). If these preferences do not satisfy basic requirements (such as transitivity), then we may speak of irrationality.

But the difficulty reappears at a different level. In a belief system, all players, rational or not, have theories–i.e., probability assessments over the types of others, and so over their actions and so on. In the foundations of decision theory, e.g. à la Savage, probabilities and utilities are derived from axioms that apply to rational players only. How, then, can an irrational player entertain probabilities–indeed, how can he have a payoff function, which is presumably derived from his utility function?

[34] Stated, but not proved, in Aumann [3, 1987].
[35] Constant throughout the belief system.

There are several ways to address this problem. One is to change the definition of a belief system, by specifying for each type whether or not it is "rational," and then specifying theories and payoff functions only for types that are specified as rational.[36] This would rescue Theorems 6.1 and 6.2, since in those results, probability assessments and payoff functions of irrational types play no role.

It would, however, not rescue Theorem 6.3, since there one needs common knowledge of the conjectures, but only mutual knowledge of rationality. Thus it appears that irrational players must entertain conjectures; for this they should be rational. To resolve this problem, one may distinguish between *subjective* and *objective* theories–and so also between subjective and objective knowledge, conjectures, and rationality. Thus, think of the common prior in Theorem 6.3 as refereeing to the "objective" assessment of a fixed outside observer Otto. Call type s_i *objectively* rational if its choice maximizes its expected payoff when calculated according to the objective conjecture (the one that Otto would hold if given s_i's information), and according to its own payoff function. This differs from *subjective* rationality of s_i (the notion of rationality hitherto used), which requires that s_i's choice be maximal according to s_i's own conjecture, not Otto's. Similarly, say that a type s_i knows some event E *objectively* if given s_i's information, Otto would ascribe probability 1 to E. Theorem 6.3 can then be understood as asking for mutual objective knowledge of the players' payoff functions and of objective rationality, and for common objective knowledge of objective conjectures. These assumptions do not demand that irrational types entertain conjectures, but they may. If they do, they are now to be thought of as subjectively–but not necessarily objectively–rational.

Needless to say, our results remain true–and almost as interesting–when rationality *is* commonly known. Thus readers who are dissatisfied with the above interpretations may simply assume common knowledge of rationality.

j. Knowledge of Equilibrium The conclusions of our theorems 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 the case of Theorems 6.2 and 6.3, though, it is in fact mutual knowledge of order 1–but not necessarily of any higher order–that σ is a Nash equilibrium. In the case of Theorem 6.1, it need not even be mutual knowledge of 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.

k. Conclusions Where does all this leave us? Are the "foundations" of Nash equilibrium more secure or less secure than previously imagined? Do our results strengthen or weaken the case for Nash equilibrium?

First, in assessing the validity of a game-theoretic solution concept, one should not place undue emphasis in its a priori rationale. At least as important is the question of how successful the concept is in providing insight in applications, how tractable it is, and, relatedly, even the extent of

[36] It would be desirable to see whether and how one can derive this kind of modified belief system from Mertens-Zamir type construction.

its aesthetic appeal. On all these counts, Nash equilibrium has proved its worth.

This said, the current results do indicate that the a priori case for Nash equilibrium is a little stronger than conventional wisdom had granted. Common knowledge turns out to play a more limited role–at least in games with two players–than previously thought. The reader may object that even mutual knowledge of, say, payoff functions is implausible; but indisputably, common knowledge of payoff functions is more so. It is true that our epistemic conditions for Nash equilibrium in games with more than two players involve common knowledge (of conjectures); indeed, it was surprising to discover that the conditions for equilibrium in the n-person case are stronger than in the two-person case. Perhaps Nash equilibrium rests on firmer foundations in two-person than in n-person games.

It should be remembered that the conditions for Nash equilibrium described here are sufficient, but not necessary. Perhaps there are other ways of looking at Nash equilibrium epistemically; if so, their nature is as yet unclear.

There also are non-epistemic ways of looking at Nash equilibrium, such as the evolutionary approach (e.g., Maynard Smith [12, 1982]). Related to this is the idea that a Nash equilibrium represents a societal norm. In the end, these viewpoints will perhaps provide a more compelling basis for Nash equilibrium than those involving games played by consciously maximizing, rational players.

Finally, the apparatus of this paper–belief systems, conjectures, knowledge, mutual knowledge, common knowledge, and the like–has an appeal that extends beyond our immediate purpose of providing epistemic conditions for Nash equilibrium. The apparatus offers a way of analyzing strategic situations that corresponds nicely to the concerns that we have all experienced in practice–what is the other person thinking, what is his true motivation, does he see the world as I do, and so on.

Appendix: Extensions and Converses

We start with some remarks on Theorem 6.3 and its proof. First, the conclusions of the Theorem 6.3 continue to hold under the slightly weaker assumption that the common prior assigns positive probability to $\boldsymbol{\varphi} = \varphi$ being commonly known, and there is a state at which $\boldsymbol{\varphi} = \varphi$ is commonly known and $\mathbf{g} = g$ and the rationality of the players are mutually known.

Second, note that the rationality assumption is not used until the end of Theorem 6.3's proof, after (6.3) is established. Thus if we assume only that there is a common prior that assigns positive probability to the conjectures φ^i being commonly known, we may conclude that all players i have the same conjecture σ_j for other players j, and that each φ^i is the product of the σ_j with $j \neq i$; that is, the $n-1$ conjectures of each player about the other players are independent.

Third, if in Theorem 6.3 we assume that the game being played is commonly (not just mutually) known, then we can conclude that also the rationality of the players is commonly known.[37] That is, we have

Proposition A1 *Suppose that at some state s, the game g and the conjectures φ^i are commonly known and rationality is mutually known. Then at s, rationality is commonly known. (Note that common priors are not assumed here.)*

Proof. Set $G := [g]$, $F := [\varphi]$, $R_j := [j \text{ is rational}]$, and $R := [\text{all players are rational}] = R_1 \cap \cdots \cap R_n$. In these terms, the proposition says that $CK\,(G \cap F) \cap K^1 R \subset CKR$. We assert that it is sufficient for this to prove

$$K^2\,(G \cap F) \cap K^1 R \subset K^2 R. \tag{A1}$$

Indeed, if we have (A1), an inductive argument using Lemma 5.5 and that $E \subset E'$ implies $K^1(E) \subset K^1\,(E')$ (which follows from Lemma 5.2) yields $K^m\,(G \cap F) \cap K^1 R \subset K^m R$ for any m; so taking intersections, $CK\,(G \cap F) \cap K^1 R \subset CKR$ follows.

Let j be a player, B_j the set of actions a_j of j to which the conjecture φ^i of some other player i assigns positive probability. Let $E_j := [\mathbf{a}_j \in B_j]$ (the event that the action chosen by j is in B_j). Since the game g and the conjectures φ^i are commonly known at s, they are a fortiori mutually known there; so by Lemma 6.1, each action in B_j maximizes g_j against φ^j. Hence $E_j \cap G \cap F \subset R_j$. At each state in F, each player other than j knows that j's action is in B_j; that is, $F \subset \bigcap_{i \neq j} K_i E_j$. So $G \cap F \subset (\bigcap_{i \neq j} K_i E_j) \cap (G \cap F)$. So Lemmas 5.2 and 5.5 yield

$$K^2\,(G \cap F) \subset K^2(\bigcap\nolimits_{i \neq j} K_i E_j) \cap K^2(G \cap F) \subset K^1(\bigcap\nolimits_{i \neq j} K_i E_j) \cap K^2(G \cap F) \subset$$
$$K^1(\bigcap\nolimits_{i \neq j} K_i E_j) \cap K^1(\bigcap\nolimits_{i \neq j} K_i (G \cap F)) = K^1(\bigcap\nolimits_{i \neq j} K_i (E_j \cap G \cap F)) \subset K^1(\bigcap\nolimits_{i \neq j} K_i R_j).$$

[37] This observation, for which we are indebted to Ben Polak, is of particular interest because in many applied contexts there is only one game under consideration, so it is of necessity commonly known.

Hence using $R \subset R_j$ and $R_j = K_j R_j$ (Corollary 5.2), we obtain

$$K^2 (G \cap F) \cap K^1 R \subset K^1(\bigcap\nolimits_{i \neq j} K_i R_j) \cap K^1 R \subset K^1(\bigcap\nolimits_{i \neq j} K_i R_j) \cap K^1 R_j = K^1(\bigcap\nolimits_{i \neq j} K_i R_j) \cap K^1 K_j R_j = K^2 R_j.$$

Since this holds for all j, Lemma 5.2 yields[38] (A.2). ■

Our fourth remark is that in both Theorems 6.2 and 6.3, mutual knowledge of rationality may be replaced by the assumption that each player knows the others to be rational; in fact, all players may themselves be irrational at the state in question. (Recall that "know" means "ascribe probability 1"; thus a player may be irrational even though another player knows that he is rational.)

We come next to the matter of converses to our theorems. We have already mentioned (at the end of Section 2) that the conditions are not necessary, in the sense that it is quite possible to have a Nash equilibrium even when they are not fulfilled. In Theorem 6.1, the action n-tuple $\mathbf{a}(s)$ at a state s may well be a Nash equilibrium even when $\mathbf{a}(s)$ is not mutually known, whether or not the players are rational. (But if the actions are mutually known at s and $\mathbf{a}(s)$ is a Nash equilibrium, then the players *are* rational at s; cf. Remark 10.2.) In Theorem 6.2, the conjectures at a state s in a two-person game may constitute a Nash equilibrium even when, at s, they are not mutually known and/or rationality is not mutually known. Similarly for Theorem 6.3.

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 Theorem 6.1, 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 6.2 and 6.3, 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 6.2 and 6.3 are met.

These "converses" show that the sufficient condition 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 Sections 7 and 8.

[38] The proof would be simpler with a formalism in which known events are true ($K_j E \subset E$). See Section 10e.

References

[1] Armbruster, W., and W. Boege, "Bayesian Game Theory," in Moeschlin, O., and D. Pallaschke (eds.), *Game Theory and Related Topics*. Amsterdam: North-Holland, 1979.

[2] Aumann, R., "Agreeing to Disagree," *Annals of Statistics*, 4, 1976, 1236-1239.

[3] Aumann, R., "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica*, 55, 1987, 1-18.

[4] Boege, W., and Th. Eisele, "On Solutions of Bayesian Games," *International Journal of Game Theory*, 8, 1979, 193-215.

[5] Brandenburger, A., and E. Dekel, "The Role of Common Knowledge Assumptions in Game Theory," in Hahn, F., (ed.), *The Economics of Missing Markets, Information, and Games*. Oxford: Oxford University Press, 1989.

[6] Geanakoplos, J., and H. Polemarchakis, "We Can't Disagree Forever," *Journal of Economic Theory*, 28, 1982, 192-200.

[7] Harsanyi, J., "Games of Incomplete Information Played by 'Bayesian' Players, I-III," *Management Science*, 14, 1967-68, 159-182, 320-334, 486-502.

[8] Harsanyi, J., "Games with Randomly Disturbed Payoffs: A New Rationale for Mixed Strategy Equilibrium Points," *International Journal of Game Theory*, 2, 1973, 1-23.

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

[10] Kreps, D., and R. Wilson, "Sequential Equilibria," *Econometrica*, 50, 1982, 863-894.

[11] Lewis, D., *Conventions: A Philosophical Study*. Cambridge: Harvard University Press, 1969.

[12] Maynard Smith, J., *Evolution and the Theory of Games*. Cambridge: Cambridge University Press, 1982.

[13] Mertens, J.-F., and S. Zamir, "Formulation of Bayesian Analysis for Games with Incomplete Information," *International Journal of Game Theory*, 14, 1985, 1-29.

[14] Myerson, R., "Refinements of the Nash Equilibrium Concept," *International Journal of Game Theory*, 7, 1978, 73-80.

[15] Nash, J., "Non-Cooperative Games," *Annals of Mathematics*, 54, 1951, 286-295.

[16] Savage, L., *The Foundations of Statistics*. New York: Wiley, 1954.

[17] Selten, R., "Spieltheoretische Behandlung eines Oligopolmodels mit Nachfragetragheit," *Zietschrift fur die gesante Staatswissenschaft*, 121, 1965, 301-324.

[18] Selten, R., "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory*, 4, 1975, 25-55.

[19] Tan, T., and S. Werlang, "The Bayesian Foundations of Solution Concepts of Games," *Journal of Economic Theory*, 45, 1988, 370-391.