Holt, Debra J.

Working Paper

Coherent Belief Revision and Equilibrium Selection in Games

Queen's Economics Department Working Paper, No. 892

Provided in Cooperation with:
Queen's University, Department of Economics (QED)

Suggested Citation: Holt, Debra J. (1993) : Coherent Belief Revision and Equilibrium Selection in Games, Queen's Economics Department Working Paper, No. 892, Queen's University, Department of Economics, Kingston (Ontario)

This Version is available at:
https://hdl.handle.net/10419/189215

Queen's Economics Department Working Paper No. 892

Coherent Belief Revision and Equilibrium Selection in Games

Debra J. Holt

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

12-1993

Discussion Paper #892

Coherent Belief Revision and Equilibrium Selection in Games

by

Debra J. Holt
Queen's University

May 1994

Coherent Belief Revision and Equilibrium Selection in Games

Debra J. Holt

Department of Economics
Queen's University
Kingston, Ontario K7L 3N6

email: holt@qed.econ.queensu.ca

Queen's University Discussion Paper #892

December 1993

I thank Bob Aumann, Daphne Koller, Bart Lipman, and participants at the Workshop on Game Theory and Computer Science, Fourth Stony Brook Summer Institute in Game Theory, and TARK '94 for valuable comments and suggestions.

ABSTRACT

What predictions about behavior in extensive form games can we rigorously justify as implied by rationality or mutual or common knowledge of that rationality? This paper models rational belief revision in extensive form games and uses numerical methods to examine the implications for equilibrium predictions. In general, in dynamic strategic choice situations, the method of belief revision adopted by the players will determine which of a set of possible equilibria (or non–Nash strategy choices) is selected. This intuition has motivated critiques of subgame perfection, including the argument that the reasoning supporting backward induction is paradoxical.

This paper examines this question with models of the dynamics of transitions between epistemic states. These models give belief revision functions which are well–defined for updating on zero probability events and are equivalent to Bayesian updating where that is defined. Thus the gap between the standard game model and the intuition driving some of the critiques of backward induction can be bridged with this approach. It appears that the class of equilibria that we obtain with this model are neither implied by, nor imply, Nash equilibrium. The sequential nature of the definition of individual rationality and the possibility of rational belief revision yield a model which implies sequential equlibrium when there is sufficient mutual knowledge of rationality and selects equilibria that satisfy forward induction. But since neither knowledge of the other players' sequential rationality nor correct beliefs are imposed, non–Nash, but individually sequentially rational, outcomes are also supported in equilibrium.

Keywords: rationality in games, belief revision, equilibrium selection.
JEL Classification Numbers: C72, C73, C63.

Coherent Belief Revision and Equilibrium Selection in Games

I. Introduction

What predictions about behavior in extensive form games can we rigorously justify as implied by rationality or mutual or common knowledge of that rationality?

This paper models rational belief revision in extensive form games and uses numerical methods to examine the implications for equilibrium predictions. Reasoning that leads to the subgame perfect equilibrium is counterfactual reasoning. For example, player I in the centipede game shown in figure 1 may reason as follows: *if I were to move across at my last information set then my opponent would move down in the subsequent move, therefore I would be better off moving down at my last information set.* A sequence of such reasoning leads to the conclusion that the subgame perfect equilibrium in the centipede game has player I moving down at his first information set. The logical problem is that if player II does get a chance to move at the second node, there are many possible ways that he could revise his belief that he will never get a chance to move.

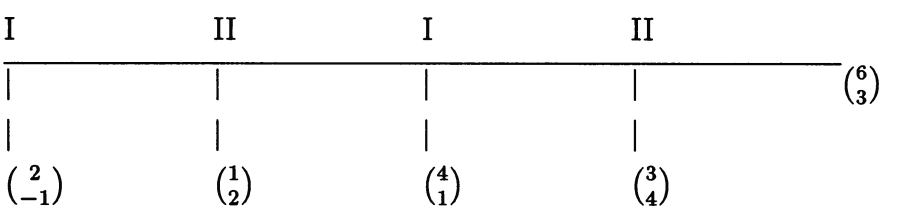

Figure 1

But depending on how player II revises his beliefs, his best action may be to play across. If so, the initial set of propositions describing player I's beliefs is not consistent with the actual belief revision process of player II. If player I knows player II's belief revision process then it may be best for player I to choose across at the first node.

In general, in dynamic strategic choice situations, the method of belief revision adopted by the players will determine which of a set of possible equilibria (or non-Nash strategy choices) is selected. This intuition has motivated critiques of subgame perfection, including the argument that the reasoning supporting backward induction is paradoxical.[1]

This paper examines this question with models of the dynamics of transitions between epistemic states. An epistemic model gives a static representation of knowledge and belief. A model of the dynamics must incorporate a model of how one rationally makes the transition from one epistemic state (roughly, a state of belief) to another. Epistemic models are homomorphic to Bayesian models (under certain axioms); that is, for a given probability ordering over states, there is an epistemic state that represents the same knowledge and belief, and there is a probability ordering (not unique) that is consistent with any epistemic state. Models of the dynamics of epistemic states yield belief revision functions which are well-defined for updating on zero probability events and are equivalent to Bayesian updating where that is defined.[2] Thus the gap between the standard game model and the intuition driving some of the critiques of backward induction can be bridged with this approach.

This approach is similar to that adopted by Aumann and Brandenberger [1991] who establish epistemic conditions sufficient for Nash equilib-

---

[1] See for example, Bicchieri [1992], Binmore [1987], and Reny [1992].

[2] See Gärdenfors [1988].

rium in normal form games. There is no need to model belief revision in normal form games; thus epistemic models are sufficient to analyze knowledge and rationality in this context. The players may use counterfactual reasoning to reason to equilibrium; but, even if the player is surprised, any belief revision that occurs is irrelevant to the analysis of the game. In their paper knowledge is defined as belief with probability one. Therefore the players can be wrong — they can "know" something that is not true.

Aumann [1993] uses an epistemic model to formulate the conditions under which, "in perfect information games, common knowledge of rationality implies backwards induction." In his formulation epistemic models are sufficient for the analysis even though he is looking at extensive form games. The dynamics of belief revision are not required because the players' strategies are implemented by agents, one for each information set. One feature of his model that differs from the Aumann and Brandenberger model and from the model in this paper is that knowledge is not equated with probability one belief; by his definition, a player, or agent, cannot know something that is not true. His main result is not only that common knowledge of rationality implies backwards induction but also that, for every game of perfect information, common knowledge of rationality is possible.

In this paper, I consider models of extensive form games in which the players take their own actions and are not irrevocably tied to any decision they may make before the beginning of the game. Also, as in Aumann and Brandenberger [1991] I define knowledge as probability one belief. A player can "know" something that is not true. Therefore, a crucial component of the model is a description of how a player can "rationally" or "coherently" move from one epistemic state to another when some evidence has been obtained that contradicts something that is known in the first instance.

I incorporate a model of belief revision developed by Gärdenfors.[3] Be-

[3] The belief revision model and its connection with other models of knowledge are explored

lief revision is a transition from one epistemic state to another. Think of representing an epistemic state by a list of all the propositions a person knows to be true in that state; this set of propositions is the person's "belief state". In this framework the set of propositions representing the player's knowledge is both complete and consistent. That is, all logical implications of the set of propositions are included; and, the set cannot contain both a proposition and its negation. A belief revision function defines a transition from one belief state to another when new information is aquired. In this paper the emphasis is on the instances in which the new information contradicts some proposition in the initial belief state.

The representation of players' knowledge or beliefs with belief states is isomorphic to a representation in epistemic logic. In fact Gärdenfors proves that belief revision functions and axioms on belief revision can be taken as the primitives and standard epistemic models generated from that base. Briefly, let K represent belief with probability one. To say that a proposition a is in person i's belief state is equivalent to saying that $K_i a$ is true. It is true of belief states that for every proposition a, exactly one of the following is true:

a is accepted in i's belief state ($K_i a$ is true)
a is rejected ($K_i - a$ is true, and $-a$ is accepted)
a is undetermined ($-K_i - a$ is true, i.e. a is possible)

Attention is restricted to epistemic logics in which $(Kp \rightarrow p)$ is not an axiom; a person can believe that p is true with probability one and actually be wrong. Axioms on positive and negative introspection do hold; that is, $(Ka \rightarrow KKa)$ is a thesis and $(-Ka \rightarrow K - Ka)$ is a thesis.[4]

in Gärdenfors [1988].

[4] The system is obtained by removing the axiom of necessity from the system S5. The properties are very similar to S5 except in the relation between truth values of propositions with and without epistemic operators. See Hughes and Cresswell [1968] for analysis of these systems.

To illustrate, return to the centipede game example where the belief state for player II might include propositions such as: *I am rational; my opponent is rational; I prefer the higher of any two payoffs*; and so on. A logical implication of one possible consistent set of propositions is that player I will move down at the first node. In this case, from completeness we know that player II's belief state includes the proposition: *Player I moves down at node 1.* Then suppose player II is called on to move at the second node. When he adds the sentence stating that play has reached this node his set of propositions is no longer consistent. In order to maintain consistency, some of the sentences that were previously held to be true must be discarded. And it may be necessary to add other sentences to obtain completeness.

There are a number of ways to remove sentences, or contract the set, and add new ones, or expand the set. Gärdenfors gives some criteria for rational belief revision such as "the contraction should be minimal". He also proposes that sentences be removed in reverse order of their "epistemic value". An ordering of epistemic value is context dependent; in games, one natural ordering would give higher epistemic value to sentences that are associated with higher expected game payoffs. Another plausible ordering would assign higher epistemic value to sentences that have greater predictive value. However, these two orderings may not be compatible.

Bicchieri [1992] shows in several examples that different belief revision functions based on different orderings of epistemic value may be associated with predictions of different equilibrium refinements. However, it is extremely cumbersome to carry out this analysis except for the simplest examples.

I use numerical methods to examine the implications of various belief revision functions on equilibrium predictions. The formal model serves as a base in which to systematically determine, for classes of games, which specific conditions on knowledge, rationality, and belief revision functions

are linked with (logically imply) various equilibrium notions. Stalnaker [1992] presents a model of normal form games in which this program can be carried out. The approach I use is a natural extension for extensive form games.

There are two reasons that it is advantageous to use belief states to represent the epistemic conditions. First, the model of belief revision that Gärdenfors develops is most naturally applied with belief states; a belief revision function is defined on belief states. Second, questions about logical implications of an epistemic state can be answered most efficiently with numerical methods when the belief state representation is used.

Hooker [1988] describes a quantitative method for logical inference; I use that method to analyze the implication for behavior in games of different assumptions on beliefs and belief revision. An inference problem in propositional logic can be written as an integer programming problem. Both the inference problem in propositional calculus and integer programming problems are known to be NP-complete. So one may reasonably ask whether it is sensible to solve the inference problem in its integer programming representation instead of directly. But in practice, many integer programming problems are relatively easy to solve; and the generalized covering models of the inference problems are among this class.[5] The evidence from applications and from experiments indicates that the integer programming representation is easier to solve than the inference problem directly.

These quantitative methods are used to carry out a systematic numerical analysis of the question: which predictions about behavior in extensive form games are logically implied by the various sets of propositions representing the players' knowledge, beliefs and rationality?

The model is introduced in the next section while the third section

---

[5] See Hooker [1988] for a thorough analysis of this approach to logical inference.

presents examples which illustrate some of the implications for equilibrium play. I conclude with a brief discussion of these preliminary results.

II. Model.

Assume we are given an extensive form game, Γ, with perfect recall and no moves by nature.[6] Such an extensive form game consists of a finite tree, a set of players, a partition of non-terminal nodes into information sets for the players, a labelling of arcs according to the action taken, and payoffs for each player for each terminal node.

Representation of Γ.

The first step in the analysis is to represent the game in propositional logic. To do so, we associate a unique atomic proposition with each arc in the game tree. A typical proposition will be a_k. The set of these is denoted $\mathcal{A}$. These are the only atomic propositions. From these, we can generate all the other propositions necessary to describe the game and rational play in the game. Since there is a one–to–one relationship between arcs and these atomic propositions, I will often use "the arc a_k" to mean the arc corresponding to atomic proposition a_k. Note that a node is completely described by the set of arcs emanating from it. So we will typically identify nodes with the appropriate set of atomic propositions.

Let $\mathcal{I} = \{I_1, \ldots, I_n\}$ be a partition of $\mathcal{A}$ such that I_i is the set of arcs controlled by a player i. A finer partition, $\mathcal{J}$, partitions each element of $\mathcal{I}$ so that each event is a set of arcs emanating from a given information set. Let A_{ij} be a typical set of arcs emanating from one information set of player i. When there are R nodes in an information set, we denote the different nodes as $A_{ij}^1, \ldots, A_{ij}^R$. When there is little danger of confusion we will refer to nodes generically as A.

Given any set A of atomic propositions representing arcs, let $C(A)$

---

[6] This is for simplicity of exposition. An extension to the case which allows for moves by nature is feasible, though perhaps not straightforward.

denote the proposition that exactly one $a_k \in A$ is true. Then for each node A, the proposition "$C(A)$" must hold; that is, a player chooses one and only one action at a node. Whenever A and A' are nodes in the same information set for player i, we know that the set of actions at each node must be the same by the definition of an information set. It follows that there is a one–to–one relationship between A and A'. Therefore we can write $A = \{a_1, \ldots, a_K\}$ and $A' = \{a'_1, \ldots, a'_K\}$ where a_k and a'_k are associated with arcs labelled by the same action. In order to respect the information structure of the game, we require "a_k if and only if a'_k," for $k = 1, \ldots, K$. We refer to these as the choice conditions on the game. Choice conditions for player i refer to the $C(A)$ for those nodes A where player i moves.

The dynamic structure of the game is described by the use of an operator, P, on atomic propositions where Pa is interpreted as "the action labelling arc a is played." For a set of arcs $A = \{a_1, \ldots, a_K\}$ let $PA = \{Pa_1, \ldots, Pa_K\}$. Each node A is reached by a unique path from the initial node, A_0, by the definition of the tree. Let $\mathcal{P}(A)$ denote the conjunction of Pa over the set of a in the path leading to A. The dynamic structure is defined by a set of propositions of the form, "$C(PA)$ if and only if $\mathcal{P}(A)$," for all nodes A. We take $\mathcal{P}(A_0)$ to always be true; that is, the initial node is always reached.

A strategy for player i is $\mathbf{a}^i \subset I_i$ such that the choice conditions for player i hold. Let $\mathbf{a}$ be a strategy profile for all players and let $\mathbf{a}^{-i}$ denote the strategies of all players except i. Let $\mathbf{A}^i$ be the set of possible strategies for player i. Define $\mathbf{A}$ and $\mathbf{A}^{-i}$ analogously. Strategies are related to the dynamic structure of the game in the following way. Given any strategy profile, $\mathbf{a}$, there is a unique terminal node z which is reached. Let $\mathcal{P}(z)$ denote the proposition that the actions leading to z are played. Then we require "$\mathbf{a} \rightarrow \mathcal{P}(z)$."

Beliefs about strategies are not necessarily correct. We denote player i's beliefs about other players' strategies by $\mathbf{b}^{-i} \in \mathbf{A}^{-i}$. A player i can have

different beliefs about strategies at different information sets; these will be determined by the belief revision functions for player i. At any information set, there will be some $\mathbf{b}^{-i}$ consistent with reaching that point in the game. For an information set $A_{ij} = \{A_{ij}^1, \ldots, A_{ij}^R\}$ let $\mathcal{P}(A_{ij})$ be the disjunction over r of $\mathcal{P}(A_{ij}^r)$. In other words, this is the statement that information set A_{ij} is reached. Let $\mathbf{a}_j^i$ be the set of actions by player i that is consistent with reaching information set A_{ij}. Then the set of feasible beliefs at that information set, denoted $\mathbf{B}_j^{-i} \subset \mathbf{A}^{-i}$, is the set $\mathbf{b}^{-i}$ such that "$\mathbf{b}^{-i}$ and $\mathbf{a}_j^i$ imply $\mathcal{P}(A_{ij})$." $\mathbf{A}_j^i$ will denote player i's set of continuation strategies from information set A_{ij}.

Note that $\mathbf{b}^{-i}$ denotes the strategy profile that player i knows is the true one — where knowledge is belief with probability one. That is, the strategy profile b^{-i} is in i's belief set but may not correspond to the objectively correct strategy profile, $\mathbf{a}^{-i}$. In particular, when we write a_k this means that the player controlling that arc intends to take that action; when we write b_k this means that another player believes the first intends to take action a_k.

Representation of rationality.

Rationality is defined sequentially for each player. Let z be a typical terminal node and denote the set of terminal nodes $\mathcal{Z}$.

The game Γ includes a payoff function for each player, $u_i : \mathcal{Z} \to \mathbf{R}$. Let $u_i(\mathbf{a})$ be the $u_i(z)$ for the terminal node z such that $\mathbf{a} \to \mathcal{P}(z)$.

Definition of rationality. Let R_i denote the proposition that, at each information set A_{ij}, for all $\mathrm{b}^{-i} \in \mathbf{B}_j^{-i}$, $\mathbf{b}^{-i}$ implies that player i's choice a_j^i satisfies $u_i(a_j^i, \mathbf{b}^{-i}) \geq u_i(\alpha_j^i, \mathbf{b}^{-i})$ for all $\alpha_j^i \in \mathbf{A}_j^i$.

III. Examples

The first example is of the four node centipede game introduced earlier in the paper under assumptions that yield the familiar result — player I moves down at the first node, there are no unanticipated events, thus there is no need to revise beliefs. The second example is of the same game but with fewer restrictions on the players' knowledge — one player knows the other's belief revision function and early play of across is supported. The third example deals with an imperfect information game and illustrates that, with rationality mutually known, the equilibrium is sequential (and the non-sequential Nash equilibrium is ruled out).

Example 1.

The game in the familiar extensive form representation is reproduced in figure 2. Labels a_k on the arcs represent the atomic propositions.

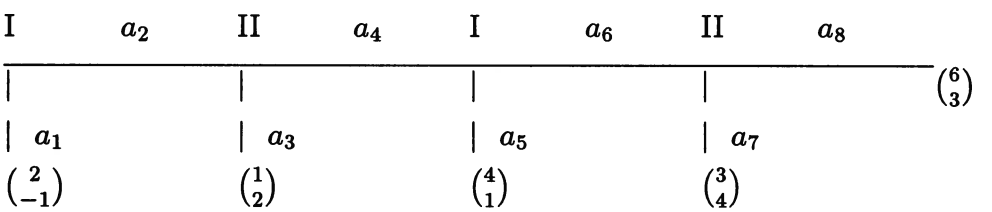

Figure 2

$$I_1 = \{a_1, a_2, a_5, a_6\}$$

$$I_2 = \{a_3, a_4, a_7, a_8\}$$

$$A_{11} = \{a_1, a_2\}$$

$$A_{12} = \{a_5, a_6\}$$

$$A_{21} = \{a_3, a_4\}$$

$$A_{22} = \{a_7, a_8\}$$

$$\mathbf{A}_1^1 = \mathbf{A}^1$$

$$\mathbf{A}_1^2 = \mathbf{A}^2$$

$$\mathbf{A}_2^1 = \{a_5, a_6\}$$

$$\mathbf{A}_2^2 = \{a_7, a_8\}$$

$$C(A_{11}) = (a_1 \vee a_2) \wedge -(a_1 \wedge a_2)$$

$$C(A_{12}) = (a_5 \vee a_6) \wedge -(a_5 \wedge a_6)$$

$$C(A_{21}) = (a_3 \vee a_4) \wedge -(a_3 \wedge a_4)$$

$$C(A_{22}) = (a_7 \vee a_8) \wedge -(a_7 \wedge a_8)$$

$$C(PA_{11})$$

$$C(PA_{21}) \text{ iff } Pa_2$$

$$C(PA_{12}) \text{ iff } (Pa_2 \wedge Pa_4)$$

$$C(PA_{22}) \text{ iff } (Pa_2 \wedge Pa_4 \wedge Pa_6)$$

The propositions above completely describe the game Γ. Next we construct propositions describing, for each player, his rationality and beliefs about the other player's rationality (but not beliefs about the other player's strategy). These propositions added to those describing Γ comprise the belief base. The belief base is just sufficient to generate the belief set; the belief set is the belief base plus its logical implications. We will denote the belief bases by $\mathcal{B}_i$.

Note that common belief (or knowledge) of rationality will itself preclude or imply certain beliefs about players' strategies. Once the belief base is specified, we first check which beliefs are consistent with each individual player's belief base. Then, we determine which beliefs are mutually consistent. Finally, we determine which paths $\mathcal{P}(\mathbf{a})$ implied by the beliefs are mutually consistent. By this I mean that information sets at which a

belief revision process is necessary will not be reached given the implied paths. The propositions on rationality are given below, by information set, where $R_i(A_{ij})$ is the conjunction of propositions defining i's rationality at information set A_{ij} and $R_i = \wedge_j R_i(A_{ij})$.

$$\begin{aligned}
R_1(A_{11}) =&((b_3, b_7) \rightarrow a_1) \wedge \\
&((b_3, b_8) \rightarrow a_1) \wedge \\
&((b_4, b_7) \rightarrow a_2) \wedge \\
&((b_4, b_8) \rightarrow a_2)
\end{aligned}$$

$$\begin{aligned}
R_1(A_{12}) =&((b_4, b_7) \rightarrow a_5) \wedge \\
&((b_4, b_8) \rightarrow a_6)
\end{aligned}$$

$$\begin{aligned}
R_2(A_{21}) =&((b_2, b_5) \rightarrow a_3) \wedge \\
&((b_2, b_6) \rightarrow a_4)
\end{aligned}$$

$$R_2(A_{22}) = ((b_2, b_6) \rightarrow a_7)$$

$$R_1 = R_1(A_{11}) \wedge R_1(A_{12})$$

$$R_2 = R_2(A_{21}) \wedge R_2(A_{22})$$

These propositions represent rationality of the two players in this game. A belief base for player I must include the propositions describing Γ and R_1 and similarly for player II. In this example, it is assumed that rationality is commonly known and that rationality of both players is true; there are no surprises. Therefore the players' belief bases are identical; both include the propositions on Γ, R_1 and R_2.

Now we want to determine which equilibria are implied by these $\mathcal{B}_1$ and $\mathcal{B}_2$. Let the proposed equilibrium strategy profile be denoted $\mathbf{a}^*$. In this example, the process is quite simple. Because we assume that the

rationality and the game are commonly known (and true), we can work with the one, common, belief base. The procedure is as follows: put the belief base in conjunctive normal form; translate the belief base into a set of constraints for an integer programming problem; minimize $\mathbf{a}^*$ subject to the constraints; if the minimized value of $\mathbf{a}^*$ is 1 then implication follows, otherwise not.

The equilibrium implied by these belief bases is $\mathbf{a}^* = (a_1, a_3, a_5, a_7)$ which is the backward induction solution.

Example 2.

The second example is of the same basic game. Rationality is mutually known; that is, each player has full belief in the rationality of the other player. Now, however, player II has a belief revision function — he knows how he would revise his beliefs if he were to be surprised by having a chance to move at A_{21}. Assume player I knows player II's belief revision function, denoted $\mathcal{C}$.

$$\mathcal{B}_1 = \{\Gamma, R_1, R_2, \Gamma \subset \mathcal{B}_2, R_1 \subset \mathcal{B}_2, \mathcal{C}\}$$

$$\mathcal{B}_2 = \{\Gamma, R_2, R_1, \Gamma \subset \mathcal{B}_1, R_2 \subset \mathcal{B}_1\}$$

Note that a_2 is inconsistent with $\Gamma \wedge R_1 \wedge R_2$ (from example 1).

Let $\mathcal{C}$ be the function which, upon observing Pa_2, removes $(R_2 \subset \mathcal{B}_1)$ from $\mathcal{B}_2$ and adds the propositions a_2 and $Pa_2 \rightarrow a_6$.

The procedure for determining the equilibrium implied by the belief bases in this example is analogous to that of the first example except that there are now two distinct belief bases. We check for implication within each belief base in the same way. Then there is an additional check that the implied $(\mathbf{a}^1, \mathbf{b}^{-1})$ and $(\mathbf{a}^2, \mathbf{b}^{-2})$ are consistent along the actual play path.

Informally, note that given C, $R_2(A_{21}) \rightarrow a_4$. Given that player I knows R_2 and C, $R_1(A_{11}) \rightarrow a_2$ and $R_1(A_{12}) \rightarrow a_5$. Then $\mathbf{a}^1 = (a_2, a_5)$, $\mathbf{b}^{-1} = (b_4, b_7)$, $\mathbf{a}^2 = (a_4, a_7)$, and $\mathbf{b}^{-2} = (b_2, b_6)$. Beliefs are inconsistent and player II is surprised by I's play of a_5 instead of a_6, but neither player is faced with the necessity of revising beliefs. Observed play is (a_2, a_4, a_5).

Example 3.

This game in this example is illustrated in figure 3. The a_k labels on the arcs represent the atomic propositions.

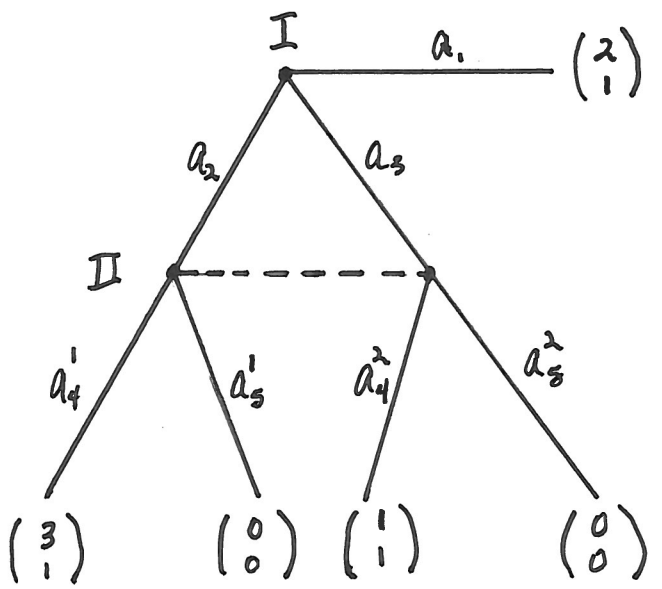

Figure 3.

$$I_1 = \{a_1, a_2, a_3\}$$

$$I_2 = \{a_4^1, a_5^1, a_4^2, a_5^2\}$$

$$A_{11} = \{a_1, a_2, a_3\}$$

$$A_{21} = \{a_4^1, a_5^1, a_4^2, a_5^2\}$$

$$A_{21}^1 = \{a_4^1, a_5^1\}$$

$$A_{21}^2 = \{a_4^2, a_5^2\}$$

$$C(A_{11}) = (a_1 \vee a_2 \vee a_3) \wedge - (a_1 \wedge a_2) \wedge -(a_1 \wedge a_3) \wedge -(a_2 \wedge a_3)$$

$$C(A_{21}) = ((a_4^1 \vee a_5^1) \wedge -(a_4^1 \wedge a_5^1)) \wedge ((a_4^2 \vee a_5^2) \wedge -(a_4^2 \wedge a_5^2)) \wedge (a_4^1 \text{ iff } a_4^2) \wedge (a_5^1 \text{ iff } a_5^2)$$

$$C(PA_{11})$$

$$C(PA_{21}) \text{ iff } (Pa_2 \vee Pa_3)$$

These propositions describe the game, Γ. Next, as before, we construct propositions describing each player's rationality according to the definition given in section II above.

$$R_1 = (b_4 \rightarrow a_2) \wedge (b_5 \rightarrow a_1)$$

$$R_2 = (b_2 \rightarrow a_4) \wedge (b_3 \rightarrow a_4)$$

Assume that rationality is mutually known. As in the previous examples, let Γ denote the set of propositions describing the game.

$$\mathcal{B}_1 = \{\Gamma, R_1, R_2, (\Gamma \subset \mathcal{B}_2)\}$$

$$\mathcal{B}_2 = \{\Gamma, R_2, R_1, (\Gamma \subset \mathcal{B}_1)\}$$

The procedure here is analogous to that used to generate the previous two examples. The only equilibrium under our definition of rationality and with mutual knowledge of the rationality is $\mathbf{a}^* = (a_2, a_4)$. Notice there are two Nash equilibria but the Nash equilibrium that is not sequential is ruled out in this model. Also note that this result goes through even without R_1 in $\mathcal{B}_2$.

Example 4.

The fourth game is illustrated in figure 4. Again, the a_k labels on the arcs represent the atomic propositions.

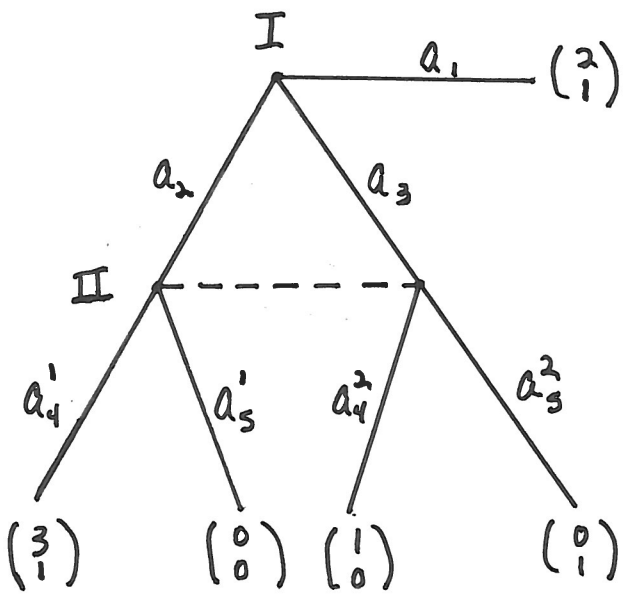

Figure 4.

The propositions which represent the game are the same as for example 3. However, because of the payoff differences to player 2, the rationality

conditions are different. The propositions describing each player's rationality are again constructed according to the definition in section II.

$$R_1 = (b_4 \rightarrow a_2) \wedge (b_5 \rightarrow a_1)$$
$$R_2 = (b_2 \rightarrow a_4) \wedge (b_3 \rightarrow a_5)$$

Again assume that rationality is mutually known and let Γ denote the set of propositions describing the game.

$$\mathcal{B}_1 = \{\Gamma, R_1, R_2, (\Gamma \subset \mathcal{B}_2)\}$$
$$\mathcal{B}_2 = \{\Gamma, R_2, R_1, (\Gamma \subset \mathcal{B}_1)\}$$

The only equilibrium implied by these belief bases is $\mathbf{a}^* = (a_2, a_4)$. In this example, there are two sequential Nash equilibria, (a_2, a_4) and (a_1, a_5), but only the first of these satisfies the forward induction criteria. This model rules out the sequential equilibrium that does not satisfy forward induction.

IV. Concluding remarks.

The examples presented here are quite simple so that the method can be clearly illustrated. Using these numerical methods, one can derive implications for more complicated games and larger classes of games. From other results currently being analyzed, it appears that the class of equilibria that we obtain with this model are neither implied by, nor imply, Nash equilibrium. The sequential nature of the definition of individual rationality and the possibility of rational belief revision yield a model which implies sequential equilibrium when there is sufficient mutual knowledge of rationality and selects equilibria that satisfy forward induction. But since neither knowledge of the other players' sequential rationality nor correct beliefs are imposed, non-Nash, but individually sequentially rational, outcomes are also supported in equilibrium.

References

Aumann, Robert J., "Backwards Induction and Common Knowledge of Rationality", 1993.

Aumann, Robert J., "Notes on Interactive Epistemology", 1992.

Aumann, Robert and Adam Brandenburger, "Epistemic Conditions for Nash Equilibrium", Harvard Business School Working Paper, 1991.

Bicchieri, Cristina, "Knowledge-dependent Games: Backwards Induction", in *Knowledge, Belief, and Strategic Interaction*, Cambridge University Press, 1992.

Binmore, Ken, "Modelling Rational Players", *Philosophy and Economics*, 1987.

Gärdenfors, Peter, *Knowledge in Flux: Modeling the Dynamics of Epistemic States*, MIT Press, 1988.

Hooker, John N., "A Quantitative Approach to Logical Inference", *Decision Support Systems*, 1988.

Hughes, G.E. and M.J. Cresswell, *An Introduction to Modal Logic*, J.W. Arrowsmith Ltd., Bristol, 1968.

Stalnaker, Robert, "Knowledge, Belief and Counterfactual Reasoning in Games", in Bicchieri & Skyrms (eds.) *Evolution, Learning and Dynamics in Games*, Cambridge University Press, forthcoming.