HYPOTHETICAL KNOWLEDGE

AND

GAMES WITH PERFECT INFORMATION*

by

Dov Samet **

December 1993

* The author wishes to thank Robert Aumann, Itzhak Gilboa, Joe Halpern, Daniel Lehmann, David Schmeidler and Robert Stalnaker for helpful comments.

** Faculty of Management, Tel Aviv University, Tel Aviv, Israel

1. Introduction

Finite games with perfect information are the simplest of games. Yet they seem to be the source of continuing debates and conflicting intuition. In particular, opinions diverge widely on the relation between rationality and the backward induction equilibrium of these games. (e.g. Aumann 1993, Basu 1990, Ben Porath 1992, Bicchieri 1989, 1994, Binmore 1987, Binmore and Brandenbuger 1990, Reny 1992.) These opinions can be roughly divide into three groups. Some claim that common knowledge of rationality is possible in such games and that it implies the backward induction equilibrium. Some maintain that it is not. Others insist that backward induction is inconsistent with common knowledge of rationality or even that common knowledge of rationality itself is a contradictory term in such games. An adequate model in which we can express formally what players do, think, believe and conjecture could enable a rigorous formulation and comparison of these diverging opinions. It appears that there are simple models that can do the job. The basic elements of such models are the following.

(1) An information structure consisting of a state space and partitions of the state space, one for each player, which define knowledge operators;

(2) An assignment of one strategy combination to each state;

and optionally,

(3) Players' probability distributions over the state space.

In such models, to which we shall refer as *standard* models, we can describe players' strategies, their knowledge and beliefs about others' strategies, their knowledge and beliefs about what others know and believe, and so on.

This paper, in a nutshell, disputes the adequacy of the standard model for the description of games with perfect information and proposes a model which is adequate for this purpose. We show that standard models fail to capture an important structural aspect of strategic thinking and therefore leave unformalized many intuitive arguments that depend on this aspect. The model we propose captures this aspect by giving a fuller and more faithful account of strategic thinking. Using this model we re-examine the relation between rationality and backward induction. In the model, we give formal expression to statements about the reasoning of players in games with perfect information – statements that cannot be formalized in the standard model.

In the next section we explain in which way the said structural aspect of strategic thinking is not represented in the standard models, and why it should be represented. For the time being, we shall explain what this structural aspect is. A strategy specifies an action in each node of the game tree. When we assign a strategy to a state, this strategy specifies the actions of a player even in nodes that are known to him not to be reached in the state. The simple intuitive meaning of this is that such an action is what the player hypothesizes he would have taken had his node been reached. This last hypothesis is a compound, structured sentence. It is formulated in terms of two clauses, one about reaching the node, the other about taking an action. This structure, the relation between the clauses and the hypothesis is absent from the standard model.

In section 3 we present an extended information structure in which the relation between sentences expressing hypothetical thinking and their components can be represented. Except for a knowledge operator, it provides for each player i a *hypothetical knowledge operator* that determines for each pair of events, H (the hypothesis) and E, the event that player i hypothesizes (or simply thinks) that if H were true, then he would know E. We define such operators axiomatically and characterize them in terms of *hypothesis transformations* on the partitions.[1]

In section 4 we define models for games with perfect information. In order that an extended information structure models a game we assign to each of its states a *play* in the game, that is, the sequence of actions taken in the game. This is a great departure from the standard model in which strategies are assigned to states. Thus the primitive notion here is purely behavioral rather than strategic. Strategies in this model are what they should be, namely, cognitive constructions. The strategy of a player in a state is determined by the hypotheses he holds about actions he would take in nodes were they reached. We show that a strategy combination in a state is uniquely defined and generates the actual play in that state.

Section 5 deals with rationality and backward induction. Here rationality is a property of behavior, and when we say that a player is rational we simply mean that his behavior is. For a given state and node we say that the player of that node is *rational at the node* if: (i) the node is reached in the given state, (ii) there is no number x such that the player knows that his action there yields him a payoff less than x, and he hypothesizes that another action would yield him a payoff of at least x. A player is *rational in a state* if he is rational in each of his nodes on the play of this state.

[1] Extended informaion structures have some resemblence to semantic models of logics of caounterfactuals. These logics were introduced in Stalnaker 1968, and further studied in Stalnaker and Thomason 1970, and Lewis 1973. Hypothetical knowledge operators differ from counterfactuals in being epistemic operators. For a comparison between the logic of counterfactuals and the logic of hypothetical knowledge see Samet 1994.

In our model common knowledge of rationality does not imply that the path is the one generated by backward induction. To see this, suppose that a certain state which is common knowledge the path is an equilibrium path other than the backward induction path. Rationality of the players in this state is determined by their behavior along the path. If players have the 'right' hypotheses there (namely, that had they deviated from the path the game would have developed according to the equilibrium strategies) then they are rational and there is common knowledge of rationality.

To explain backward induction in terms of rationality, in our model, we need hypotheses about hypotheses about ...rationality rather than common knowledge of it. For this purpose we define *common hypothesis* which is constructed iteratively, like common knowledge, only that hypothetical knowledge operators are iterated rather than knowledge ones. Common hypothesis has, though, more structure than common knowledge. *Common hypothesis that if node v is reached then event E holds true*, is the event obtained by iterating knowledge operators where the sequence of hypotheses is the sequence of events corresponding to the nodes leading to v. The iteration ends with the event E. We say that *node rationality is common hypothesis* if for each node v it is common hypothesis that if v is reached, then the behavior of the player at v is rational there. We prove in section 5 the following theorems. The first theorem gives epistemic conditions that suffice to guarantee that the backward induction play results.

Theorem 5.3. *If there is common hypothesis of node rationality, then players play the backward induction path.*

Note that the provision in Theorem 5.3, means that for each node there is common hypothesis that the player in that node plays rationally *there*. It does not mean that for each node there is common hypothesis that the player in that node is rational. The next theorem shows that almost the opposite is true.

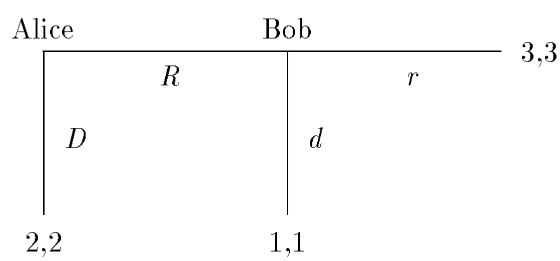
Theorem 5.5. *If there is common hypothesis of node rationality (which implies by the previous theorem that players play the backward induction path), then for each node which is not on the backward induction play there is common hypothesis that if it were to be reached, then it would be the case that not all players are rational.*

These theorems formalize frequently heard arguments concerning backward induction. Such a formalization is impossible in the standard model.

To simplify the presentation we did not include in the model elements that could give a fuller account of players' reasoning and behavior. Thus, time is absent from the model — we analyze the game at a point in time before it is played. We do not introduce into the model beliefs (which would be given by probability distributions over the state space). We do not foresee any difficulties in replicating the results for games with beliefs. Finally, we studied only games with perfect information, but we think that extended information structures are the appropriate tool for the investigation of games in extensive form in general.

2. What standard models lack

The shortcomings of standard models are exhibited even in very simple games. To demonstrate these shortcomings we use the following game which is used in Aumann 1993, as an illustration.



Alice and Bob has each, two strategies in this game; S_R and S_D for Alice; S_r and S_d for Bob. The intuitive meaning of strategies equates them with certain sentences. In Bob's case:

S_d = 'If Bob's node is reached he plays d'

S_r = 'If Bob's node is reached he plays r'

To say that Bob's strategy in a state ω is S_d means simply that it is true in ω that if Bob's node is reached he plays d. Equivalently, in terms of events, Bob's strategy in a state ω is S_d if ω is in the event[2] $[S_d]$.

The sentences S_d and S_r are compound sentences. For example, S_d has two component sentences 'Bob's node is reached' and ' Bob plays d'. Can S_d be constructed from these component sentences? Rephrasing the question in terms of events; can we determine the event $[S_d]$ from the events [Bob's node is reached] and [Bob plays d]? We show that in standard models this is

[2] For a sentence S we denote by $[S]$ the event that S is true, i.e., the set of all states in which S is true.

impossible. That is, the event $[S_d]$ is not constructed from the component events. Consider the following standard model.

> **Model 1.** There is only one state in this model, which is therefore common knowledge. The strategies in this state are S_D and S_d.

In this model, the event $[S_d]$ consists of the single state ω of the model, while the event $[S_r]$ is the empty set. Bob's node is not reached in ω and he plays neither d nor r. Thus the component events of both events $[S_d]$ and $[S_r]$ are the same – the empty set. Yet $[S_d]$ and $[S_r]$ are different events. Hence these events cannot be constructed from their components[3]. Thus, the states in which S_d is true are determine arbitrarily without consulting the component sentences. The strategy S_d is considered as a primitive notion in the standard model.

The intention of the previous discussion is by no means to claim that sentences like 'If Bob's node were to be reached, then he would have played d', have no meaning. Just the opposite. Such sentences describe hypothetical thinking which is essential to decision and game theory. We only claim that the standard model does not lend such sentences any meaning since it takes them as primitive. There is nothing wrong with this since every theory must have some primitive notions (such are also actions and payoffs in the standard model). We show however that by taking strategy as a primitive notion we fail to formalize the most basic ideas that relate behavior to rationality. Consider the following model.

> **Model 2.** There is only one state in this model, which is therefore common knowledge. The strategies in this state are S_R and S_r.

Both Model 1 and 2, describe a Nash equilibrium. We can easily explain the different behavior in these equilibria in terms of rationality. In Model 1, Alice knows that if Bob's node were to be reached he would have played, irrationally, d, and being herself rational she plays D. In Model 2, Alice knows that if Bob's node were to be reached he would have played, rationally, r, and being herself rational she plays R.

This explanation (or any other one which involves Bob's rationality) cannot be expressed in the standard model. Bob's rationality, in Model 1 for example, is refered to in the component sentence 'Bob plays irrationally d'. But the specification of Bob's strategy S_d in this model is not

[3] To conform with English grammar, the sentence describing strategy S_d, in Model 1, should be 'If Bob's node were to be reached, then he would have played d', since it is known in ω that Bob's node will not be reached.

related, as we argued before, to this component sentence or the corresponding event. All that can be said in Model 1 is that Alice knows that Bob's strategy is S_d. To conclude from this that she knows Bob would behave irrationally had he been reached cannot be represented in the model.

Now the strategies in Model 2 are the backward induction strategies while those in Model 1 constitute another equilibrium. Thus we cannot account for the difference between these two equilibria in terms of events that depend on Bob's rationality, let alone common knowledge of rationality.

There is a way to overcome this obstacle, and express the difference between the equilibria, in the standard model, in terms of rationality. To this end we have to adopt a notion of rationality which is radically different from the one we were using. Instead of describing actions as being rational or not we ascribe such properties to strategies. Thus, instead of saying that had Bob's node been reached he would have chosen, irrationally, d, we say that choosing strategy S_d, in Model 1, is irrational on Bob's part. With this definition of rationality, the event [Bob is irrational] is just $\{\omega\}$ and of course Alice knows that Bob is irrational. A generalization of this type of rationality is used in Aumann 1993, to prove that common knowledge of rationality implies backward induction.

The difference between the two notions of rationality is the following. According to the first notion it is always an action actually taken by the player, that is, behavior, which is judged to be rational or irrational. Thus in Model 1, where Bob dose not take any action he cannot be considered irrational according to this notion of rationality, which is the reason why we cannot relate the equilibrium to Bob's irrationality. The second notion applies rationality to strategies which are primitive entities that cannot be expressed in terms of actual actions (in standard models) and can be nonbehavioral and payoff irrelevant (as in Model 1).

In summary, the standard models are inadequate to express the outcome of the game in terms of player's rationality, when rationality is the elementary behavioral notion of rationality.

This is not the only drawback of standard models. Similarly to strategy sentences, all sentence which express hypothetical thinking cannot be represented in terms of their components, in these models. Consider for example the following sentence S.

$S =$ 'Player i (who chooses action a) thinks that had she chosen action b, player j would have known she, player i, is irrational.'

Such sentences are usually heard and debated when backward induction is discussed, usually as an argument against it's justification. Unfortunately we cannot test the validity of arguments concerning S, using the standard model, since S cannot be formalized in this model. That is, there is no way we can define the event $[S]$ in terms of the events [Player i chooses b] and [Player j knows that player i is irrational], for reasons similar to the ones given above. We could have determined arbitrarily the states in which S is true, much the same way we did for strategies, but this does seem to be fruitful, since there must be some non-arbitrary relations between the event $[S]$ and events describing strategies.

The conclusion from the above discussion is that an adequate model for games in extensive form must provide a way to interpret as events sentences describing hypothetical thinking in terms of their components. In the next section we extend information structures to facilitate such interpretation.

3. Extended information structures

Information structures. An *information structure* for a set of players I, is a list

$$(\Omega,\ (\Pi_i)_{i\in I}),$$

where Ω is the set of *states*, and for each player i, Π_i is a partition of Ω. Subsets of Ω are called *events*. The set of all events is denoted by Σ. The set theoretic operations of complementation, intersection and union are considered as the interpretation of 'not', 'and' (conjunction) and 'or' (disjunction) respectively. The complement of event E is denoted by $\neg E$. We define a binary operator '$\rightarrow$', by $E \rightarrow F = (\neg E) \cup F$, for any two events E and F. The operator '$\rightarrow$' interprets material implication ('if...,then...').

For each state ω, we denote by $\Pi_i(\omega)$, the unique element of Π_i containing ω. Using the partitions we define knowledge operators as follows. For each i, the *knowledge operator* $K_i : \Sigma \rightarrow \Sigma$ is defined by,

$$K_i(E) = \{\omega \mid \Pi_i(\omega) \subseteq E\},$$

for each event E. We usually omit the parentheses and write K_iE. The event K_iE is the event that i *knows* E. For each player i define an operator L_i by $L_iE = \neg K_i \neg E$, for each event E. The event L_iE is the event that i considers E *possible*. Clearly $L_iE = \{\omega \mid \Pi_i(\omega) \cap E \neq \emptyset\}$. The 'all know' operator, K_I, is defined by $K_IE = \cap_{i\in I} K_iE$. The event that E is *common knowledge* is $\cap_{n\geq 1}(K_I)^n E$.

Hypothetical knowledge. We want to enrich information structures so that hypothetical knowledge can be represented by events. For that purpose we define for each player i, a binary operator

$K_i(\cdot,\cdot)$ such that for any two events $H \neq \emptyset$ and E, $K_i(H,E)$ is the event that i hypothesizes that he would have known E had the hypothesis H been true. We write $K_i^H(E)$ or $K_i^H E$ instead of $K_i(H,E)$.

The following definition provides a list of axioms that we want the hypothetical knowledge operator to satisfy. For the rest of this section we suppress the subscript i in K_i, K_i^H L_i,and Π_i.

Definition 3.1. *An operator*

$$K\colon (\Sigma \setminus \{\emptyset\}) \times \Sigma \to \Sigma,$$

is called a hypothetical knowledge operator if it satisfies for all events $H \neq \emptyset$ and E, in Σ, the following axioms:

(K1) $K^H E = K K^H E$

(K2) $K^H E = K^H K E$,

(K3) $\neg K^H E = K^H \neg K E$,

(K4) *For any family of events* $\{E_\alpha\}$, $\cap_\alpha K^H E_\alpha = K^H \cap_\alpha E_\alpha$,

(K5) $K^H L H = \Omega$,

(K6) $LH \cap K^H E = LH \cap KE$.

The intuition behind these axioms is as follows. We think of $K^H E$ as representing a hypothesis of the player about his knowledge in a hypothetical situation H. Thus, there are two states of the player's mind which are involved. One is the actual state of mind that this hypothesis is part of which. The other is a hypothetical state of the player's mind which is referred to by his hypothesis. To produce a statement 'Had H been the case, I would have known E' the player has to probe, being in his actual state of mind, a hypothetical state of mind of his, one that would have prevailed had H been the case. We make the following two assumptions. (i) The hypothetical state of mind is not a special type of state of mind; it enjoys standard properties of states of mind. The adjective 'hypothetical' only defines its relation to the actual state of mind. (ii) The hypothetical state of mind can be fully examined by the player in his actual state of mind. That is, he can tell precisely what is known and what is not known to him in this state of mind.

In light of this interpretation of hypothetical knowledge the axioms are rather natural. The first axiom (K1) is expressing the idea that $K^H E$ is a hypothesis made by the player in his actual state of mind and as such is known to him. In other words, we consider the statement 'Had H

been the case, I would have known E' as having the same meaning as 'I know that had H been the case, I would have known E'.

In hypothetical states of mind, as in all states of mind in a partition model, the player knows E if and only of he knows that he knows E. If this hypothetical state of mind is fully open to the player's investigation, then stating 'Had H been the case, I would have known E' must be true precisely when 'Had H been the case, I would have known that I know E' is true. This is axiom (K2).

Axiom (K3) reflects also the full inseparability of the hypothetical state of mind by the player. For each E, he either finds that in his hypothetical state of mind he knows E or he finds there that he does not know it (and in the latter case he knows that he does not know it). Thus the player would deny that if H had been the case he would have known E, precisely when he finds that in this hypothetical situation he would have not known E (and therefore know that he would have not known it). This is what (K3) says.

Axiom (K4) states that hypothetical states of mind have a standard property of states of mind. That is, knowing hypothetically that certain events are simultaneously true is the same as simultaneously knowing hypothetically that each one of them holds.

The following two axioms relate the hypothetical state of mind to the hypothesis H. It is always the case, no matter what the hypothesis H is, says (K5), that one can state 'Had H been the case I would have known it is possible (that is, I wouldn't have known it is false)'.

Axiom (K6) deals with the limit case in which the hypothesis H does not deserve this title because it is indeed considered possible by the player[4] (in his actual state of mind). In this case the player does not have to go too far to find his hypothetical state of mind. There is no reason (and indeed it will be strange) to assume that if indeed H were the case his state of mind will be any different from his actual state of mind in which he considers H possible. But then, hypothetical knowledge is precisely actual knowledge, which is what axiom (K6) states.

Hypothesis transformations. In the previous subsection we formalized hypothetical knowledge by an operator on events. In formulating the axioms this operator should satisfy, we made no use of the structure of the state space.

[4] The English grammatical form 'Had H been true. . . ' is used when the hypothesis H is known to be false. Otherwise, 'If H is true . . . ' is used. The hypothetical knowledge operator covers both forms and has no grammatical equivalent in English.

We start with an informal discussion which motivates the definition that follows. In the previous subsection we did not formalize the notions of actual states of mind and hypothetical ones. Rather we used these notions as intuitive guide lines in formalizing the properties of hypothetical knowledge operator. Now we formalize actual and hypothetical states of mind, and the relation between them, in terms of the agents' partitions and then show how this formalization is related to hypothetical knowledge operator.

The player state of mind in a state ω can be identified with his partition element $P = \Pi(\omega)$. This event describes completely every thing that the player knows. Suppose now that P is the actual state of the player's mind. To find what he knows when a hypothesis H holds true, the player must have in mind a hypothetical state of mind which is another element in his partition. Let's denote this element by $T(P, H)$.

Obviously, since $T(P, H)$ is the player's state of mind when H is true it should be the case that the two events intersect non-trivially, because otherwise the player knows in $T(P, H)$ the H is false. In addition, if P intersect non-trivially H then the player considers H possible and therefore it is reasonable to assume that $T(P, H)$ is simply H.

This leads us to the following definition.

Definition 3.3. *A hypothesis transformation on* Π *is a map* $T \colon \Pi \times (\Sigma \setminus \{\emptyset\}) \to \Pi$ *that satisfies the following two conditions,*

(T1) $T(P, H) \cap H \neq \emptyset$,

(T2) if $P \cap H \neq \emptyset$, then $T(P, H) = P$.

The following theorem shows that hypothetical knowledge operators are those operators that can be represented by hypothesis transformations.

Theorem 3.4. *Let* (Ω, Π) *be an information structure with set of events* Σ*, and* K *a binary operator* $K \colon (\Sigma \setminus \{\emptyset\}) \times \Sigma \to \Sigma$*. Then* K *is a hypothetical knowledge operator iff there exists a hypothesis transformation* T*, on* Π*, such that,*

$$K^H E = \cup_{P \in \Pi} \{P \mid T(P, H) \subseteq E\}. \tag{3.5}$$

Moreover, for a given hypothetical knowledge operator there exists a unique hypothesis transformation T *that satisfy* (3.5).

Definition 3.5. *An extended information structure for a set of players I is a list,*

$$\big(\Omega,\ (\Pi_i, T_i)_{i\in I}\big),$$

where Ω is the set of states and for each player i, Π_i is a partition of Ω, and T_i is a hypothesis transformation on Π_i.

In the axioms defining the hypothetical knowledge operator we use the knowledge operator K. But by axiom (K7), for each E, $K^{\Omega}E = KE$. Thus the hypothetical knowledge operator is an extension of the knowledge operator. Let (K1*)-(K6*) be the axioms obtained from (K1)-(K6) by replacing each K in the latter by K^{Ω}. Add also the axiom,

(K7*) $K^{\Omega}E \subseteq E$.

Axioms (K1*)-(K7*) can be used now to axiomatize both the knowledge operator and the hypothetical knowledge operator, as follows.

Theorem 3.6. *Let Ω be a state space with set of events Σ and K a binary operator $K\colon (\Sigma\backslash\emptyset)\times\Sigma \to \Sigma$. Then K satisfies the axioms (K1*)-(K7*) iff there exists an extended information structure $\big(\Omega,\ (\Pi, T)\big)$, such that the operator $K^{\Omega}(\cdot)$ is the knowledge operator corresponding to the partition Π, and K is the hypothetical knowledge operator defined by the hypothesis transformation T.*

4. Models for games with perfect information

A *game* G with perfect information consists of a finite set of players I and a finite tree with set of of non-terminal nodes (or vertices) V, a set of terminal nodes Z and root r. For each $z \in Z$ there is a unique path called *play* leading to it from the root r. With some abuse of language we call z a play when no confusion arises. For two nodes v and u we write $v \preceq u$ or $u \succeq v$ when u is a node that follows v, i.e., when u is a node in the subtree the root of which is v. The set of non-terminal nodes, V, is partitioned into subsets $(V_i)_{i\in N}$, where V_i is the set of i's nodes. For $v \in V$ we denote by $A(v)$ the set $\{a \mid (v,a)$ is an arc of the tree$\}$. If $v \in V_i$, then the nodes in $A(v)$ are called i's *actions* at v. Player i's *payoff function* is a real-valued function $h_i : Z \to R$.

A *strategy for player* i is a function $s_i\colon V_i \to V \cup Z$, such that for each $v \in V_i$, $s_i(v) \in A(v)$. A *strategy* is a list of strategies $s = (s_i)_{i\in I}$. The set of all strategies of i is denoted by S_i and the set of all strategies is denoted by S. We denote by $s(v)$ the terminal node that is reached by strategy

s from node v. Thus, if v is a terminal node, then $s(v) = v$. If $s(a)$ is defined for all $a \in A(v)$ and $v \in V_i$, then $s(v) = s\big(s_i(v)\big)$. The play that results from playing strategy s is obviously $s(r)$.

In the next definition we use extended information structures as models for games by introducing a function $\zeta\colon \Omega \to Z$. We use the following notation for events that are defined in terms of ζ. The event that node v is (or to be) reached, $\{\omega \mid \zeta(\omega) \succeq v\}$, is denoted for short by $[v]$. Note that $[v] = \cup_{a \in A(v)}[a]$, and $[a] \cap [a'] = \emptyset$ for any $[a] \neq [a']$ in $A(v)$.

Definition 4.1. *A model for a game with perfect information G with set of players I is a pair, $(\mathcal{S}, \zeta)$, where $\mathcal{S}$ is an extended information structure $\big(\Omega,\ (\Pi_i, T_i)_{i \in I}\big)$, and ζ is a map $\zeta\colon \Omega \to Z$, onto Z, such that for each player i, node $v \in V_i$ and action $a \in A(v)$,*

$$[a] \subseteq K_i\big([v] \to [a]\big). \tag{4.2}$$

The map ζ specifies for each state the behavioral aspects of the state, namely what players do in this state (i.e., which terminal node they reach). Using temporal terminology, ζ shows how the game evolves in time. We intend the extended information structure to describe the epistemic status of the players before the game is started. It tells us what the players know or hypothesize in the beginning of the game about how the game will evolve in time. Note that since ζ maps Ω on Z, for each node v, the event $[v]$ is not empty, and therefore can serve as a hypothesis for the hypothesis transformations T_i.

To understand condition (4.2) note that $[v] \to [a]$ is the event that if v is reached, then action a is taken there. Thus $K_i([v] \to [a])$ is the event that the player (who takes this action) knows that if v is reached he takes there action a, or simply, he knows that he takes a at v. Now (4.2) says that if action a is indeed to be taken by the player at node v, then the player knows it. Observe however, that it is not true that a player who plays a in some state must know it. That is, we do not require that $[a] \subseteq K_i([a])$. This is so, because in the beginning of the game, i may not know for sure that his node v will be reached.

A strategy map, which we define shortly, assigns to each state a strategy. It is constructed in our model using the hypotheses players have about their actions in all possible and all hypothetical nodes. In order to define strategy maps we need the following proposition.

Proposition 4.3. *For each player i and node $v \in V_i$, there exists a unique action a, of player i, which i hypothesizes that he would know to be his action at v, if v were to be reached. That is,*

$$\bigcup_{a \in A(v)} K_i^{[v]}\big([v] \to [a]\big) = \Omega,$$

and the events in this union are disjoint in pairs.

The uniqueness of the hypothesized actions, guaranteed by Proposition 4.3, enables us to define a strategy in each state.

Definition 4.4. *Player i's strategy map is the function $\sigma_i\colon \Omega \to S_i$, such that $\sigma_i(\omega)$ is the strategy that assigns for each $v \in V_i$ the unique action a for which $\omega \in K_i^{[v]}\big([v] \to [a]\big)$. The strategy map $\sigma\colon \Omega \to S$ is defined by $\sigma(\omega) = \big(\sigma_i(\omega)\big)_{i\in I}$.*

We denote the event that i's strategy is s_i (that is, $\{\omega \mid \sigma_i(\omega) = s_i\}$) by $[s_i]$, and the event that the strategy is s (i.e.,$\{\omega \mid \sigma(\omega) = s\}$) is denoted by $[s]$. Clearly

(4.5)
$$\begin{aligned} [s_i] &= \bigcap_{v\in V_i} K_i^{[v]}\big([v] \to [s_i(v)]\big) \\ [s] &= \bigcap_{i\in I} [s_i]. \end{aligned}$$

The following proposition shows that the strategy map is related to behavior and knowledge in an appropriate way.

Proposition 4.6. *For each strategy $s = (s_i)_{i\in I}$, player i, and state ω, if the strategy in ω is s, then the play in ω is the play generated by s and each player knows his strategy. That is,*

$$[s] \subseteq [s(r)],$$

and

$$[s_i] = K_i[s_i].$$

5. Rationality and backward induction

We assume in this section that for each player i the payoff function h_i is *nondegenerate* that is, it attains each value only once. The payoff maps η_i on Ω are defined for each player i by, $\eta_i(\omega) = h_i\big(\zeta(\omega)\big)$. The event $\{\omega \mid \eta_i(\omega) < x\}$ is denoted by $[\eta_i < x]$. The event $[\eta_i \geq x]$ is similarly defined.

Since the payoff functions are nondegenerate it is easy to show by induction that there is a unique strategy $\beta = (\beta_i)_{i\in I}$ which satisfies for each i and $v \in V_i$ the condition $\beta_i(v) = \mathrm{argmax}_{a\in A(v)} h_i\big(\beta(a)\big)$. We call β and $\beta(r)$ the *backward induction* strategy and path correspondingly.

Definition 5.1. *The event that player i is rational in node $v \in V_i$ is defined as:*

$$R(v) = [v] \cap \bigcap_{x} \bigcap_{a \in A(v)} \neg\Big(K_i([v] \rightarrow [\eta_i < x]) \cap K_i^{[a]}([a] \rightarrow [\eta_i \geq x])\Big),$$

The event that player i is rational is:

$$R_i = \bigcap_{v \in V_i} \big([v] \rightarrow R(v)\big).$$

That is, i is rational at v if this node is reached and i does not know of any bound x on his payoff, given that v is reached, that could be hypothetically surpassed by using some action a. Note that rationality in node v is a behavioral concept since it applies only to states in which v is indeed reached. Clearly, $\omega \in R_i$ iff for each node v on the play $\zeta(\omega)$, i is rational at v. Thus the rationality of i is also a behavioral concept.

We want to find sufficient conditions in terms of rationality that guarantee the backward induction play. Common knowledge of rationality does not suffice. For example, if z is any equilibrium play, then it is possible to construct a state ω such that $\zeta(\omega) = z$, $\{\omega\}$ is common knowledge and all players are rational in ω.

The following condition is stronger than the assumption of common knowledge of rationality but it is still too weak. The event that *i hypothesizes he is rational in all his nodes* is $\cap_{v \in V_i} K_i^{[v]}\big([v] \rightarrow R(v)\big)$. Common knowledge that each player hypothesizes he is rational in all his nodes is not enough to guarantee the backward induction play. The reason is as follows.

Suppose that i is the player who plays at the root r. His rationality at r depends on his hypotheses of what would happen in each action he might choose. Thus the important events to look at are of the form $K_i^{[a]}E$, where a is an action at r. Now the common knowledge assumption tells us that i knows that the player who plays at a, say j, hypothesizes he is rational at a. That is $K_i\Big(K_j^{[a]}([a] \rightarrow R_j(a))\Big)$ is true. But there is nothing that bounds i to hypothesize that if a is reached, j would behave rationally. Common knowledge of the players' hypotheses impose no inter-personal consistency on these hypotheses. Therefore what i knows about the hypotheses entertained by other players tells us nothing about what he hypothesizes their behavior or hypotheses would be. This consideration leads us to the following required condition. What we need is hypotheses about hypotheses about hypotheses... as we define next.

Definition 5.2. *For each pair of non-terminal nodes u and v such that $v \succeq u$, and event E, the event that there is common hypothesis that if u is to be reached, then E is true if v is to be reached,*

$\mathcal{H}(u, v, E)$, *is defined inductively as follows.*

$$\mathcal{H}(v, v, E) = E.$$

If $u \in V_i$, $v \succeq u$, *and* $\mathcal{H}(a, v, E)$ *is defined for the (unique) node* a *in* $A(u)$ *on the path from* u *to* v *(i.e.,* $v \succeq a$*), then*

$$\mathcal{H}(u, v, E) = K_i^{[a]}\big([a] \to \mathcal{H}(a, v, E)\big).$$

The event that there is common hypothesis of node rationality is,

$$\bigcap_{v \in V} \mathcal{H}\big(r, v, R(v)\big).$$

Theorem 5.3. *If there is common hypothesis of node rationality, then players play the backward induction path, that is,*

$$\bigcap_{v \in V} \mathcal{H}\big(r, v, R(v)\big) \subseteq [\beta(r)].$$

By the definition of common hypothesis it follows that common hypothesis of node rationality depends only on the hypotheses of the root player. The reason that the hypotheses of one player determine the play of the game is as follows. Clearly, only the hypotheses of the root player, and nothing else, determine his action at the root. What he hypothesizes about the consequents of his actual action at the root is knowledge (since the antecedent is true) and moreover, since the root player knows his action the consequents are indeed true. But among these consequents is the hypotheses of the next player about what is true if his node is reached. Since the next player's node is indeed reached, his hypotheses are knowledge and so on.

Theorem 5.3 is of interest only if in some models of the game G, the event that there is common hypothesis of node rationality is not empty. The following theorem guarantees the existence of such a model.

Theorem 5.4. *For each game with perfect information there exists a model in which*

$$\bigcap_{v \in V} \mathcal{H}\big(r, v, R(v)\big) \neq \emptyset.$$

Common hypothesis of node rationality is not common hypothesis of rationality of the players. Almost the opposite is true. Common hypothesis of node rationality implies that there is common hypothesis that not all players are rational in unreached nodes.

Theorem 5.5. *Let $R = \cap_{i\in I} R_i$ be the event that all players are rational. Then,*

$$\bigcap_{v\in V} \mathcal{H}\big(r, v, R(v)\big) \subseteq \bigcap_{u\not\leq\beta(r)} \mathcal{H}\big(r, u, \neg R\big).$$

6. Proofs

We first record several properties of hypothesis operators that we use in the sequel.

Proposition 6.1. *For all events H, E and F:*

(a) *If $E \subseteq F$, then $K^H E \subseteq K^H F$ (monotonicity).*

(b) $K^H\Omega = \Omega$.

(c) $K^H\emptyset = \emptyset$.

(d) *For any family of events $\{F_\alpha\}$, $\cup_\alpha K^H(E \to KF_\alpha) = K^H(E \to \cup_\alpha KF_\alpha)$.*

(e) $H \cap K^H E = H \cap KE$.

Proof:

(a) Using (K4), $K^H E = K^H(E \cap F) = K^H(E) \cap K^H(F) \subseteq K^H(F)$.

(b) Follows from (a) and (K5).

(c) Using (K3) and (b), $K^H\emptyset = K^H\neg K\Omega = \neg K^H\Omega = \emptyset$.

(d) It is easy to see that the knowledge operator K satisfies,

$$\cup_\alpha K(E \to KF_\alpha) = K(E \to \cup_\alpha KF_\alpha).$$

Taking the complement of $\cup_\alpha K^H(E \to KF_\alpha)$ and using (K3) and (K4) we conclude that,

$$\begin{aligned}\neg \cup_\alpha K^H(E \to KF_\alpha) &= \cap_\alpha K^H\neg K(E \to KF_\alpha)\\ &= K^H \cap_\alpha \neg K(E \to KF_\alpha)\\ &= K^H\neg \cup_\alpha K(E \to KF_\alpha)\\ &= K^H\neg K(E \to \cup_\alpha KF_\alpha)\\ &= \neg K^H(E \to \cup_\alpha KF_\alpha).\end{aligned}$$

(e) Intersect both sides of (K6) with H and note that $H \subseteq LH$. ∎

Proof of Theorem 3.4: We denote by $\mathcal{F}_i$ the range, $K_i\Sigma$, of K_i. It is closed under complementation and intersection (of any number of events) and for any event E, $K_iE = E$, if and only if, $E \in \mathcal{F}_i$.

Let K be a hypothetical knowledge operator. Consider event $H \neq \emptyset$ in Σ and $P \in \Pi$. Let $\{E_\alpha\}$ be the family of all events for which $P \subseteq K^H E_\alpha$. This family is not empty by (K5). Thus $P \subseteq \cap_\alpha K^H E_\alpha$. Let $T(P,H) = \cap_\alpha C_\alpha$, then by (K4),

$$P \subseteq K^H T(P,H). \tag{6.2}$$

By (6.2) and monotonicity we conclude that for any $R \in \Sigma$,

$$P \subseteq K^H R \quad \text{iff} \quad T(P,H) \subseteq R. \tag{6.3}$$

By (K1), $K^H R \in \mathcal{F}$ and therefore $P \not\subseteq K^H R$ iff $P \subseteq \neg K^H R$. Hence (6.3) is equivalent to,

$$P \subseteq \neg K^H R \quad \text{iff} \quad T(P,H) \not\subseteq R. \tag{6.4}$$

By (6.2) $K^H T(P,H) \neq \emptyset$ and thus by Proposition 6.1 (c), $T(P,H) \neq \emptyset$. We show now that $T(P,A) \in \Pi$. First we prove that $T(P,H) \in \mathcal{F}$. This is true if $KT(P,H) = T(P,H)$ or, by the definition of $T(P,H)$ and (K4), if $\cap_\alpha KE_\alpha = \cap_\alpha E_\alpha$. Clearly $\cap_\alpha KE_\alpha \subseteq \cap_\alpha E_\alpha$. The inverse inclusion holds if whenever $P \subseteq K^H E_\alpha$, then $P \subseteq K^H KE_\alpha$ which is true by (K2). Suppose to the contrary that $T(P,A) \notin \Pi$. Then, $T(P,H)$ is a union of more than one elements of Π, and hence there is $S \in \Pi$ such that $T(P,H) \not\subseteq S$ and $T(P,H) \not\subseteq \neg S$. It follows by (6.4) that $P \subseteq \neg K^H S$ and $P \subseteq \neg K^H \neg S$. Since $KS = S$ it follows by (K3) that $P \subseteq K^H S$ and $P \subseteq K^H \neg S$. By (K4) and Proposition 6.1 (c), $P \subseteq K^H \emptyset = \emptyset$ which is a contradiction.

To prove (T1) note that by (K5), $P \subseteq K^H LH$ which implies by (6.3) that $T(P,H) \subseteq LH$, i.e., $T(P,H) \cap H \neq \emptyset$.

To prove (T2) suppose that $P \cap H \neq \emptyset$ i.e., $P \subseteq LH$. Combining this inclusion with (6.2) we have, $P \subseteq K^H T(P,H) \cap LH$. It follows by (K6) that $P \subseteq KT(P,H) \cap LH$. But since $T(P,H) \in \Pi$, $KT(P,H) = T(P,H)$ and thus $P = T(P,H)$.

since $K^H E \in \mathcal{F}$, (3.5) holds by (6.3).

To see that T is unique, suppose T' is a hypothesis transformation satisfying (3.5). By applying (3.5) to T' we conclude that for all event P and $H \neq \emptyset$, $P \subseteq K^H T'(P,H)$. Hence by

(6.3), $T(P,H) \subseteq T'(P,H)$ and since, by the definition of hypothesis transformations, both events are in Π, $T(P,A) = T'(P,A)$.

It is straightforward to show that if K is defined by (3.5), then it satisfies (K1)–(K6), i.e., it is a hypothetical knowledge operator. ∎

Proof of Proposition 3.6: Suppose K satisfies axioms (K1*)-(K7*). By substituting Ω for H in (K2*),(K3*),(K4*) and (K7*) we have the following properties of $K^\Omega(\cdot)$. For each $E \in \Sigma$, $K^\Omega E = K^\Omega K^\Omega E$, $\neg K^\Omega E = K^\Omega \neg K^\Omega E$, and for any family of events $\{E_\alpha\}$, $\cap_\alpha(K^\Omega E_\alpha) = K^\Omega(\cap_\alpha E_\alpha)$. These properties, with (K7*), show that K^Ω is a knowledge operator and therefore there exists a partition Π on Ω which represents K^Ω. By Theorem 3.4 there is a hypothesis transformation on Π representing the binary operator K. The opposite direction of the theorem can be easily checked. ∎

Proof of Proposition 4.3: We write in this proof K for K_i. Unions and intersections with respect to a are taken over all actions $a \in A(v)$. We prove first that a is taken by i precisely when v is reached and i knows that he is playing a at v. That is,

$$[a] = [v] \cap K\big([v] \to [a]\big). \tag{6.5}$$

To see this, observe that since $[a] \subseteq [v]$, then (4.2) implies,

$$[a] \subseteq [v] \cap K\big([v] \to [a]\big). \tag{6.6}$$

Therefore,

$$[v] = \cup_a [a] = \cup_a \Big([v] \cap K\big([v] \to [a]\big)\Big). \tag{6.7}$$

By (6.6) and (6.7), it is enough to show that the events in the last union in (6.7) are disjoint in pairs. Indeed, for $[a] \neq [a']$ in $A(v)$,

$$\begin{aligned}
[v] \cap K\big([v] \to [a]\big) \cap K\big([v] \to [a']\big) &= [v] \cap K\big([v] \to ([a] \cap [a']\big) \\
&= [v] \cap K\big([v] \to \emptyset\big) \\
&= [v] \cap K\big(\neg[v]\big) \\
&= \emptyset.
\end{aligned}$$

Using (6.5) we have

$$\cup_a K^{[v]}\big([v] \to [a]\big) = \cup_a K^{[v]}\big([v] \to [v] \cap K\big([v] \to [a]\big)\big) = \cup_a K^{[v]}\big([v] \to K([v] \to [a])\big).$$

By proposition 6.1 (d), this last event is $K^{[v]}\big([v] \to \cup_a K([v] \to [a])\big)$. But by (6.7), $[v] \subseteq \cup_a K\big([v] \to [a]\big)$ and thus $[v] \to \cup_a K\big([v] \to [a]\big) = \Omega$. Therefore, by proposition 6.1 (b), $K^{[v]}\big([v] \to \cup_a K([v] \to [a])\big) = \Omega$, and we have proved the equality of the proposition.

To show that the events are disjoint in pairs, observe that by (K4), for $[a] \neq [a']$ in $A(v)$,

$$K^{[v]}\big([v] \to [a]\big) \cap K^{[v]}\big([v] \to [a']\big) = K^{[v]}\big([v] \to [a] \cap [a']\big) = K^{[v]}\big([v] \to \emptyset]\big) = K^{[v]}(\neg[v]).$$

By (K2), (K5),(K4) and proposition 6.1 (c),

$$K^{[v]}(\neg[v]) = K^{[v]}(K\neg[v]) = K^{[v]}(K\neg[v]) \cap K^{[v]}(\neg K\neg[v]) = K^{[v]}(\emptyset) = \emptyset.$$

∎

Proof of Proposition 4.6: We prove by induction that for each node v on the play leading to $s(r)$, (i.e., $v \preceq s(r)$), $[s] \subseteq [v]$. For $v = r$ this is true since $[r] = \Omega$. Suppose we have proved for $v \in V_i$ on this play, and let $a = s_i(v)$. We must show that $[s] \subseteq [a]$. By (4.5), $[s] \subseteq K_i^{[v]}\big([v] \to [a]\big)$ and therefore, by the induction assumption, $[s] \subseteq [v] \cap K_i^{[v]}\big([v] \to [a]\big)$. By Proposition 6.1 (e), $[v] \cap K_i^{[v]}\big([v] \to [a]\big) = [v] \cap K_i\big([v] \to [a]\big)$. But the last event is $[a]$ by (6.5).

The equality $[s_i] = K_i[s_i]$ follows from (4.5) and (K1). ∎

To facilitate a proof of Theorem 5.3 by induction we provide in the next lemma an inductive expression for common hypothesis.

Lemma 6.8. *For each $u \in V_i$,*

$$\bigcap_{v \succeq u} \mathcal{H}\big(u, v, R(v)\big) = R(u) \cap \bigcap_{a \in A(u)} K_i^{[a]}\big([a] \to \cap_{v \succeq a} \mathcal{H}\big(a, v, R(v)\big)\big).$$

Proof: By the definition of $\mathcal{H}$,

$$\begin{aligned}\bigcap_{v \succeq u} \mathcal{H}\big(u, v, R(v)\big) &= R(u) \cap \bigcap_{v \succ u} \mathcal{H}\big(u, v, R(v)\big)\\ &= R(u) \cap \bigcap_{a \in A(u)} \bigcap_{v \succeq a} K_i^{[a]}\big([a] \to \mathcal{H}\big(a, v, R(v)\big)\big).\end{aligned}$$

The lemma follows now by applying (K4) to the last event. ∎

To prove the theorem we use the following lemma that states that if i's node u is reached and he hypothesizes that each of his actions leads to the backward induction node that follows the action, then he knows that he ends up in one of these nodes. We denote

$$H(u,\beta) = \bigcap_{a \in A(u)} K_i^{[a]}\big([a] \to [\beta(a)]\big).$$

Lemma 6.9. *For each player i and node $u \in V_i$,*

$$[u] \cap H(u,\beta) \subseteq \bigcup_{a \in A(u)} K_i\big([u] \to [\beta(a)]\big).$$

Proof: Intersections and unions with respect to a are taken over all $a \in A(u)$. Substituting $\cup_a [a]$ for $[u]$ and using Proposition 6.1 (e), we have,

$$\begin{aligned}
[u] \cap H(u,\beta) &\subseteq \Big(\cup_a [a]\Big) \cap \Big(\cap_a K_i^{[a]}\big([a] \to [\beta(a)]\big)\Big) \\
&\subseteq \cup_a \Big([a] \cap K_i^{[a]}\big([a] \to [\beta(a)]\big)\Big). \\
&\subseteq \cup_a K_i\big([a] \to [\beta(a)]\big).
\end{aligned} \tag{6.10}$$

By (6.7) $[u] \subseteq \cup_a K_i\big([u] \to [a]\big)$. Combining this with (6.10) yields,

$$\begin{aligned}
[u] \cap H(u,\beta) &\subseteq \Big(\cup_a K_i\big([u] \to [a]\big)\Big) \cap \Big(\cap_a K_i\big([a] \to [\beta(a)]\big)\Big) \\
&\subseteq \cup_a \Big(K_i\big([u] \to [a]\big) \cap K_i\big([a] \to [\beta(u)]\big)\Big) \\
&\subseteq \cup_a K_i\big([u] \to [\beta(u)]\big),
\end{aligned}$$

which completes the proof. ■

The next Lemma claims that if the player at u is rational and he hypothesizes that each of his actions leads to the backward induction node that follows the action, then the play ends in the backward induction that follows u.

Lemma 6.11. *For each $u \in V$,*

$$R(u) \cap H(u,\beta) \subseteq [\beta(u)].$$

Proof: Let $x = h_i\big(\beta(u)\big)$. Then $[\beta_i(a)] \subseteq [\eta_i < x]$ for each $a \in A(u)$ such that $a \neq \beta_i(u)$, and $[\beta(\beta_i(u))] \subseteq [\eta_i \geq x]$. From this we can conclude that, for $a \neq \beta_i(u)$, $K_i\big([u] \to [\beta_i(a)]\big) \cap H(u,\beta) \cap R(u) = \emptyset$, and $K_i\big([u] \to [\beta(\beta_i(u))]\big) \cap H(u,\beta) \cap R(u) \subseteq [\beta(u)]$. Using Lemma 6.9 we have,

$$\begin{aligned}
R(u) \cap H(u,\beta) &= R(u) \cap [u] \cap H(u,\beta) \\
&\subseteq R(u) \cap \Big(\cup_a K_i\big([u] \to [\beta(a)]\big)\Big). \\
&\subseteq [\beta(u)].
\end{aligned}$$

∎

Proof of Theorem 5.3: We prove by induction that for each node $u \in V$,

$$\bigcap_{v \succeq u} \mathcal{H}\big(u, v, R(v)\big) \subseteq [\beta(u)].$$

If $A(u) \subseteq Z$, then $\cap_{v \succeq u} \mathcal{H}\big(u, v, R(v)\big) = R(u) \subseteq [\beta(u)]$. Suppose we have proved this for all $a \in A(u)$. Combining Lemma 6.8, Lemma 6.11 and the induction assumption we have,

$$\begin{aligned} \bigcap_{v \succeq u} \mathcal{H}\big(u, v, R(v)\big) &= R(u) \cap \bigcap_{a \in A(u)} K_i^{[a]}\big([a] \to \bigcap_{v \succeq a} \mathcal{H}\big(a, v, R(v)\big)\big) \\ &\subseteq R(u) \cap \bigcap_{a \in A(u)} K_i^{[a]}\big([a] \to [\beta(a)]\big) \\ &\subseteq [\beta(u)]. \end{aligned}$$

∎

Proof of Theorem 5.4: We define an extended information structure as follows. The state space Ω is the set of terminal nodes Z (and thus for each $z \in Z$, $\{z\} = [z]$). Each player i's partition is the finest, i.e., $\Pi_i(z) = \{z\}$ for each $z \in Z$. Define T_i for each z and $H \neq \emptyset$ as follows. If $z \notin H$, then $T_i(\{t\}, H) = \{\beta(v)\}$ for some v which is minimal (with respect to $\preceq$) in H. If $z \in H$, then $T_i(\{z\}, H) = \{z\}$. It is easy to see that T_i is a hypothesis transformation. Note that for each $u \in V_i$ and $a \in A(u)$, $T_i(\{\beta(u)\}, [a]) = \{\beta(a)\}$ and therefore,

$$\{\beta(u)\} \subseteq \bigcap_{a \in A(u)} K_i^{[a]}\big([a] \to \{\beta(a)\}\big).$$

It is easy to see that from this inclusion it follows that for each $u \in V$,

$$\{\beta(u)\} \subseteq R(u).$$

To show that $\cap_v \mathcal{H}\big(r, v, R(v)\big)$ is not empty we prove by induction that for each $u \in V$, $\{\beta(u)\} \subseteq \cap_{v \succeq u} \mathcal{H}\big(u, v, R(v)\big)$. If $A(u) \subseteq Z$ this inclusion is reduces to $\{\beta(u)\} \subseteq R(u)$ which clearly holds. Suppose we have proved this for all $a \in A(u)$, then using Lemma 6.8, monotonicity and the induction assumption we have,

$$\begin{aligned} \bigcap_{v \succeq u} \mathcal{H}\big(u, v, R(v)\big) &= R(u) \cap \bigcap_{a \in A(u)} K_i^{[a]}\big([a] \to \cap_{v \succeq a} \mathcal{H}\big(a, v, R(u)\big)\big) \\ &\supseteq R(u) \cap \bigcap_{a \in A(u)} K_i^{[a]}\big([a] \to [\beta(a)]\big) \\ &\supseteq [\beta(u)]. \end{aligned}$$

∎

References

Aumann, R., 1993; Backward Induction and Common Knowledge of Rationality, a manuscript.

Basu, K., 1990; On the Non-existence of a Rationality Definition for Extensive Games, *International Journal of Game Theory,* **19**, pp. 33–44.

Ben Porath, E., 1992; Rationality, Nash Equilibrium and Backward Induction in Perfect Information Games, The Sackler Institute of Economic Studies, Tel Aviv University, Working Paper 14–92.

Bicchieri, C., 1989; Self Refuting Theories of Strategic Interaction: A Paradox of Common Knowledge, *Erkenntnis* **76**, pp. 69–85.

Bicchieri, C., 1994; *Rationality and Coordination,* Cambridge University Press, Cambridge.

Binmore, K., 1987; Modeling Rational Players I, *Economics and Philosophy,* **3**, pp. 179–214.

Binmore, K., and A., Brandenbuger, 1990; Common Knowledge and Game Theory, in *Essays on the Foundation of Game Theory,* Oxford: Basil Blackwell, pp. 105–150.

Lewis, D., 1973; *Counterfactuals,* Harvard University Press, Cambridge.

Samet, D., 1994; The Logic of Hypothetical Knowledge, a manuscript.

Stalnaker, R., 1968; A theory of Conditionals, in N. Rescher (ed.), *Studies in Logical Theory,* Basil Blackwell, Oxford, pp, 98–112

Stalnaker, R., and R., Thomason, 1970; A semantic Analysis of Conditional Logic, *Theoria,* **36**, 246–281

Reny, P., 1992; Rationality in Extensive Form Games, *Journal of Economic Perspectives,* **6,** pp.92–100.