TEL-AVIV

14-92

The Sackler Institute of Economic Studies

מכון סאקלר לכלכלה
אוניברסיטת תל-אביב

RATIONALITY, NASH EQUILIBRIUM AND BACKWARD INDUCTION IN PERFECT INFORMATION GAMES*

by

Elchanan Ben–Porath**

Working Paper No.14–92

December, 1992

* I thank Bob Aumman, Itzhak Gilboa, Daniel Kahneman, Michael Kahneman, Ehud Kalai, Ariel Rubinstein and Dov Samet for valuable discussions and comments. I am especially indebted to my friend Eddie Dekel who has followed this work closely at all its different stages and has made many valuable suggestions and comments.

** Department of Economics, Tel–Aviv University and Department of Managerial Economics and Decision Sciences, J.L. Kellogg Graduate School of Management, Northwestern University.

THE SACKLER INSTITUTE OF ECONOMIC STUDIES
Faculty of Social Sciences
Tel–Aviv University, Ramat Aviv, Israel.

ABSTRACT

We say that a player is certain of an event A if he gives A probability 1. There is common certainty (CC) of A if the event A occurred, each player is certain of A, each player is certain that every other player is certain of A and so forth. It is shown that in a generic perfect information game the set of outcomes that are consistent with common certainty of rationality (CCR) at the beginning of the game coincides with the set of outcomes that survive one deletion of weakly dominated strategies and then iterative deletion of strongly dominated strategies. Thus, the backward induction outcome is not the only outcome that is consistent with CCR. In particular, cooperation in Rosenthal's [1981] centepede game, and fighting in Selten's [1978] chain–store game are consistent with CCR at the beginning of the game. Intuitively, the problem with the backward induction argument, an argument which seems to follow from CCR, is that it assumes that players reason according to a certain logic even at vertices which are inconsistent with that logic (i.e. vertices that would not have been reached had the players followed the backward induction logic). Next, it is shown that if, in addition to CCR, there is CC that each player gives a positive probability to the true strategies and beliefs of the other players, and if there is CC of the support of the beliefs of each player, then the outcome of the game is a Nash equilibrium outcome. The paper concludes with an extension of the model which allows for the possibility of mistakes in the implementation of the strategies (as opposed to mistakes in the reasoning of the players). It is shown that if, in addition to CCR, there is CC that there is a small probability of a mistake at every vertex, then the players choose the backward induction strategies.

I. INTRODUCTION

Let G be a finite n–person extensive–form game with perfect information. Our first interest is in a characterization of the set of outcomes that are consistent with common certainty of rationality (CCR) at the beginning of the game. Rationality means that when a player faces a decision under uncertainty, such as choosing an action when there is uncertainty about the strategies of the other players, he picks an action which maximizes his expected utility with respect to his subjective probability over the events of which he is uncertain. CCR means that each player is rational, each player gives probability 1 to the event that all the others are rational, each player gives probability 1 to the event where all the other players give probability 1 to the event that everyone is rational, and so forth.

The difference between assuming CCR and assuming an equilibrium solution is that in an equilibrium theory the assumption is that each player chooses a best response to the profile of strategies that is played by the other players. There are many situations in which it would be difficult to assume that a player has a correct expectation about the strategies of the other plays. In particular, it is easy to construct simple examples in which there is a mismatch between what player I does and what player II believes player I does, and yet each one of them is rational and furthermore there is CCR.

In a simultaneous–move game the set of outcomes that is consistent with CCR is the set of rationalizable outcomes (Bernheim (1984) and Pearce (1984).) If we assume that the beliefs of a player about what the others players are going to do can be correlated, as I assume here, then the set of outcomes which is consistent with CCR is the set of outcomes that survive iterative deletion of strongly dominated strategies. The difference between a dynamic game and a simultaneous game is that in a dynamic game the beliefs of a player can change as the game progresses. In particular, player i might find out during the game that player j is playing a strategy to which player i gave initially a zero probability. In such a case, we assume that player i picks a new belief which is consistent with the history of the game.[1] According to the definition of rationality discussed earlier, a strategy is rational only if it is a best response to the beliefs in each vertex. One approach to incorporating rationality in extensive–form games would be to use an iterative procedure, analogous to iterated deletion of strongly dominated strategies. Pearce (1984) defined rationalizability for extensive–form games in this way. For perfect information games, rationalizability coincides with backward induction. We now demonstrate by means of a simple example that although it seems that there is a compelling argument that shows that backward induction follows from CCR, this argument is actually problematic. The discussion of the example is intuitive and its basic points have already been recognized (see Reny (1988) and Gul (1989).) The purpose of the discussion is to provide some further motivation to the problem that I consider and to give an intuition for the results.

Consider the game that is discribed in figure 1. (The game is a four–stage version of the famous centepede game that was studied by Rosenthal (1981).)

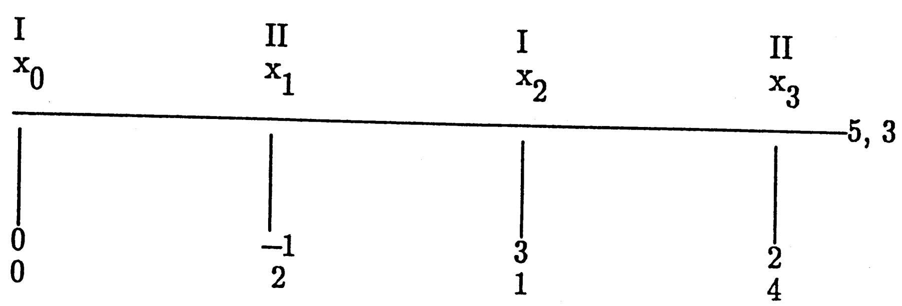

Figure 1

[1] As we will see, even when each player gives a positive probability to the profile of strategies that is actually played by the other players, the fact that player i will change his beliefs if player j deviates to a strategy to which player i gives a zero probability is important.

If player II is rational, then clearly he should play bottom at x_3.

Therefore, player I plays bottom at x_2 if the following conditions hold:

1. Player I is rational;
2. Player I believes (at x_2) that player II is rational.

Therefore, Player II plays bottom at x_1 if the following is satisfied:

1. Player II is rational;
2. Player II believes (at x_1) that: (i) Player I is rational; (ii) Player I believes (at x_2) that player II is rational.

Continuing in this manner one additional step we obtain that player I plays bottom at x_0 (which is the backward induction outcome). The only assumptions that were made in this derivation were of the type: Player i believes that player j believes... that player k is rational. So it seems that CCR implies the backward induction outcome. But, now, assume that in contrast to the above conclusion player I plays right at the beginning. What should player II do? The backward induction logic suggests that player II should play bottom because if player II plays right player I is expected to play bottom. But is it obvious that player I will play bottom at x_2? Player I plays bottom at x_2 if he thinks according to the backward induction logic, but if he thinks according to that logic why did he play to x_1? We will try to argue now that it is not at all clear how player II should interpret a move by player I to x_1. Consider the following three possible theories about player I:[2]

[2]There are many other possible theories but those listed above are sufficient to make the point.

1. Player I is not rational, with probability 0.5 he plays right and with probability 0.5 bottom.
2. Player I is rational, but he believes that player II is not rational and that player II plays with probability 0.5 right and with probability 0.5 bottom.
3. Player I is rational and believes that player II is rational. However, player I believes that at x_1 player II will adopt theory 1 as an explanation for player I's behavior.

It is easy to check that if player II believes in the first or second theory he should in fact play right in x_1. (For example, if player II believes in theory 1 then his expected utility from playing right is $0.5 \cdot 1 + 0.5 \cdot 4 = 2.5$. This is greater than his utility from playing bottom).The third theory is the only one among the three which is consistent with the backward induction prescription. Note that all these theories contain a chain of beliefs that end with the proposition that one of the players is not rational. This is not a coincidence. Reny (1988) has shown that it is impossible to have CCR at the vertex x_1. I take the point of view that once CCR collapses it is difficult to make the case that one theory is clearly more plausible than the others. If it is not clear how to interpret a move of player I to x_1, then it is not clear what player II should do, and therefore it might be reasonable for player I to move to x_1 (for example, if he believes that player II will believe in theory 1). As we will see, a move to x_1 is in fact consistent with CCR at x_0. Thus, it is possible for a game to begin with CCR and terminate without it.

This discussion emphasizes the need for a formal model in which the implications of assumptions such as CCR can be examined in a rigorous way. I propose a model which extends to dynamic games the decision–theoretic approach that was taken in the work of

Aumman(1976 and 1978), Brandenburger and Dekel (1987), Tan and Werlang (1988), and Aumman and Brandeburger (1991).[3] This model gives a complete description of the beliefs of each player at each state of the world and at each vertex in the game tree. Thus, in a given state we can talk about what player i believes at a vertex x about player j's beliefs at a vertex y ... about what player k plays at a vertex w.

Our first result, theorem 1, is a characterization of the set of outcomes that are consistent with CCR at the beginning of the game. It is shown that in a generic perfect information game this set coincides with the set of outcomes that survives one round of deletion of weakly dominated strategies and then iterative deletion of strongly dominated strategies. So, for example, in the game that is described in figure 1 only the outcomes corresponding to the payoffs (2,4) and (5,3) are not consistent with CCR at the beginning of the game. Thus, the backward induction outcome (which is also the only Nash equilibrium outcome in this game) is not the only outcome that is consistent with CCR. Next we ask, is there a natural strengthening of the CCR assumption that will imply the backward induction outcome or a Nash equilibrium outcome? Of course, we are interested in assumptions that are, in a sense, weaker than the equilibrium assumption that a player knows the strategies of the other players. Theorem 2 shows that if in addition to CCR there is, first, CC that each player gives a positive probability to the true profile of strategies and beliefs of the other players, and second, CC of the support of the beliefs of each player, then the outcome of the game is a Nash equilibrium outcome. (However, the profile of strategies that is played is not necessarily a Nash equilibrium, since the strategies may differ from a Nash equilibrium off the equilibrium path). These assumptions have a

[3] Brandenburger (1991) provides an excellent survey of the decision–theoretic approach to the analysis of normal–form games.

natural interpretation if we hypothesize that a player does not give a zero probability to an event unless there is compelling evidence, or reason, to exclude it. Under this hypothesis the first assumption is that a player does not receive compelling evidence that an event did not occur when it actually did. The second assumption is that each player is aware of compelling evidence that another player receives. These assumptions are weaker than the standard equilibrium assumptions in the sense that we do not assume that each player knows precisely the strategies or beliefs of the other players.[4] In fact in a generic normal–form game these assumptions do not restrict the set of outcomes any further than assuming only CCR.

The question now is whether there exists an attractive set of assumptions that selects the backward induction outcome. I was unable to find such a set in a model where players make no mistakes. In section 5 I consider an extension of the model which allows for the possibility of mistakes in the implementation of the strategies (as opposed to mistakes in the reasoning of the players). It is shown, (theorem 3), that if in addition to CCR there is CC, at the beginning of the game, that at every vertex there is a "small" probability for mistakes then the players choose the backward induction strategies. This result follows because a move to a vertex which is inconsistent with CCR in a world without mistakes, is now attributed to a mistake in the implementation of the strategy. Therefore CCR is maintained at every vertex and the backward induction process is valid. The observation that mistakes can have an important impact on the outcome of a game is due, of course, to Selten (1975). Selten studied mistakes in the context of an equilibrium theory and developed the notion of perfect equilibrium. Borgers (1991) introduces mistakes

---

[4] We say that our assumptions are weaker only in a sense because we have to assume common certainty and not just certainty.

into a rationalizability theory. Borgers considers general extensive–form games, he examines different assumptions on the mistakes and proposes several solution notions. The goal in the current study is to present a formal framework in which a proposition such as theorem 3 can be stated and proved in a rigorous way.

Let D(G) denote the set of outcomes that survive one round of deletion of weakly dominated strategies and then iterative deletion of strongly dominated strategies. There are three other papers in which D(G) is obtained as a solution set. Borgers (1990) and Brandenburger (1990) consider games in the normal form. Borgers derives D(G) by assuming that there is approximate common knowledge of rationality and that each player assigns a strictly positive probability to all the strategies of the other players. Brandenburger shows that Borger's derivation can be replicated by assuming common knowledge of rationality with lexicographic beliefs. Dekel and Fudenburg (1990) consider extensive form games and derive D(G) as the result of iteration of weakly dominated strategies when each player has a "small" uncertainty on the payoffs of the others. In the current study it is assumed that the payoffs are common knowledge, but I do not assume iteration of weakly dominated strategies. Such an iteration suffers from the problems that were described in our discussion of the centepede game. These problems are also discussed in Batigalli (1990), Bicierri (1988), Binmore (1987), Gilboa (1989), Gul (1989) and Reny (1988). Some of these papers contain characterizations of backward induction. However, these characterizations rely on strong assumptions on the players' beliefs in vertices that are inconsistent with CCR. In the main part of the current study sections 2–4, we are interested in characterizations of solution sets which do not put any restrictions on the beliefs of players at vertices which are inconsistent with CCR.

To summarize, the paper provides a decision–theoretic model for the analysis of extensive–form games.[5] For perfect information games this paper has two types of results: First, the set of outcomes that are consistent with CCR is characterized, and second, we examine what additional assumptions are needed to obtain traditional solution notions such as Nash equilibrium and backward induction. The paper is organized as follows: Section 2 describes the model. In Section 3 we characterize the set of outcomes that are consistent with CCR. Section 4 obtains a characterization of Nash equilibrium outcomes and in Section 5 the backward induction outcome is derived in an extension of the model which allows for mistakes.

2. THE MODEL

Let G be an n–person extensive–form game with perfect information. We use the following notation:

$N \equiv \{1,\dots,n\}$	– the set of players
X	– the set of vertices in the game tree
X_i	– the set of vertices where player i moves
x_0	– the initial vertex

[5] The model that is proposed refers only to perfect–information games. However, it can be easily extended to general extensive–form games.

$\ell: X \setminus \{x_0\} \to X$ — the precedence function, $\ell(x)$ is the predecessor of x. (The function ℓ defines the edges in the game tree.)

Z — the set of terminal vertices

$u_i: Z \to \mathbb{R}$ — the utility function for player i

S_i — the set of pure strategies for player i. (A strategy s_i for player i is a function $s_i: X_i \to X$ such that $\ell(s_i(x)) = x$)

$S \equiv \pi^n_{j=1} S_i$ — the set of profiles of strategies

$S_{-i} \equiv \pi_{j \neq i} S_j$

Throughout we will restrict attention to games in which all the payoffs are different, i.e. $\forall\, i \in N$ and $\forall z, w \in Z$ s.t. $z \neq w$, $u_i(z) \neq u_i(w)$. This class of games is generic.

A vertex y is a <u>successor</u> of a vertex x if $\ell(y) = x$. A <u>path</u> p is a sequence of vertices $x_1,\ldots,x_m$ such that x_{i+1} is a successor of x_i. A profile of strategies $s \in S$ determines a path $p(s)$ from x_0 to a terminal vertex $z(s)$,($z(s)$ will also be denoted as $z(p)$, where $p = p(s)$.) A profile s is <u>consistent</u> with a vertex x, if x is on $p(s)$. A profile s is <u>consistent</u> with a path p if s is consistent with every vertex on p. $s_i \in S_i$($s_{-i} \in S_{-i}$) is <u>consistent</u> with $x \in X$ (with a path p) if there exists $s'_{-i} \in S_{-i}$ ($s'_i \in S_i$) such that (s_i, s'_{-i}) ((s'_i, s_{-i})) is consistent with x (with p). We let $S_i(x)$ ($S_{-i}(x)$) denote the set of strategies for player i (the set of profile of strategies for the other players) that are consistent with x. $X(s_i)$ denotes the set of vertices that are consistent with s_i. The <u>height</u> of a vertex x, $h(x)$, is defined as follows: $h(x) = 0$ for $x \in Z$, $h(\ell(x)) = h(x) + 1$. The <u>height</u> of a game G, $h(G)$, is the height of its initial vertex.

Let $G(x)$ denote the subgame of G that is obtained by taking x as the initial vertex. We let $v_i(x)$ denote the maxmin payoff for player i in the game $G(x)$

$$v_i(x) \equiv \max_{s_i \in S_i(x)} \min_{s_{-i} \in S_{-i}(x)} u_i[z(s_i, s_{-i})].$$

It is well known that in a perfect information game the maxmin and minmax in pure strategies are equal so that

$$v_i(x) = \min_{s_{-i} \in S_{-i}(x)} \max_{s_i \in S(x)} u_i[z(s_i; s_{-i})].$$

Also, we can choose $s_{-i} \in S_{-i}(x)$ such that s_{-i} is the minmax strategy in every subgame.

The situation that we want to analyze is an interaction with asymmetric information. Following the standard approach of modelling such interactions we let T denote the set of possible states of the world. A state $t \in T$ specifies all the relevant aspects of the interaction of which at least one of the players is uncertain. The assumption is that the set T is common knowledge among the players. In our model each player is uncertain of the strategies and beliefs of the other players. Thus, a state $t \in T$ is an n–tuple $t = (t_1,\ldots,t_n)$ where t_i specifies the strategy and the beliefs of player i on the true state of the world. t_i will be called the type of player i. In a dynamic game the beliefs of a player will typically change as the game progresses. So a type t_i specifies a pair (s_i, μ_i) where $s_i \in S_i$ and μ_i: $X \to \Delta(T)$. In the sequel we will identify t_i with the pair (s_i, μ_i) which it specifies. Define:

$$T_i \equiv \{t_i \mid \exists t \in T \text{ s.t. } t = (t_1,\dots,t_i,\dots,t_n)\}.$$

T_i is the set of possible types of player i. We can assume w.l.o.g. that $T = \Pi_{i=1}^{n} T_i$. To avoid tedious technicalities I will assume that T is finite.[6] We assume that each player knows his own type but has no prior information about the type of another player. We can therefore now write μ_i as $\mu_i : X \to \Delta(T_{-i})$ where $T_{-i} = \Pi_{j \neq i} T_j$. We let $s(t)$ ($s_i(t_i)$) denote the profile of strategies (the strategy for player i) that is played in the state t. A state t is consistent with a vertex x if x is consistent with $s(t)$. We say that $t_{-i} \in T_{-i}$ is consistent with $x \in X$ if there exists a type $t_i \in T_i$ such that (t_i, t_{-i}) is consistent with x. Define:

$$T_{-i}(x) \equiv \{t_{-i} \mid t_{-i} \in T_{-i} \text{ is consistent with } x\}.$$

We say that a type $t_i = (s_i, \mu_i)$ is Bayesian rational if:

1. For every $x \in X$, $\mu_i(x)$ gives probability 1 to profiles that are consistent with x. (Formally, $\mu_i(x)[T_{-i}(x)] = 1$).
2. Beliefs are updated according to Bayes Law whenever that is possible. Formally, suppose $x, y \in X$ and x preceeds y. Let $A \subseteq T_{-i}$ be a set. If
$\mu_i(x)[T_{-i}(y)] > 0$ then
$$\mu_i(y)[A] = \frac{\mu_i(x)[A \cap T_{-i}(y)]}{\mu_i(x)[T_{-i}(y)]}.$$

---

[6]Although I have not proven it I am confident that with the appropriate definitions this assumption is innocuous.

3. The strategy s_i is a best response to the beliefs μ_i, i.e. for every $x \in X$ s_i maximizes the expected utility of player i in the subgame $G(x)$ w.r.t. the beliefs $\mu_i(x)$.

A strategy $s_i \in S_i$ is <u>rational</u> if there exists a rational type $t_i \in T_i$ such that $s_i = s_i(t_i)$.

We now turn to a formal definition of common certainty – CC – of an event $A \subseteq T$ at a state t at a vertex x. First, we need some additional notation.

Let $\overline{T} \subseteq T$ be a set of states, $\overline{T}_i \equiv \{t_i \mid \exists\, t' \in \overline{T} \text{ s.t. } t_i' = t_i\}$.

$\mu_i(t_{-i} \mid t_i, x)$ denotes the probability that type t_i of player i gives to the profile $t_{-i} \in T_{-i}$, at the vertex x. We let $\mu_i(t_j \mid t_i, x)$ denote the probability that t_i gives, at the vertex x, for player j being of type t_j.

Define $T^m(t,x)$, $m = 0,1,\ldots$ as follows:

$T^0(t,x) \equiv \{t\}$

For $m \geq 1$

$T^m(t,x) \equiv \{t' \mid \text{for some } i \in N \;\; t_i' \in T_i^{m-1}(t,x) \text{ and } \mu_i(t_{-i}' \mid t_i', x) > 0\}$

$T^1(t,x)$ is the minimal event of which every player is certain at (t,x), $T^2(t,x)$ is the minimal event E such that each player is certain, at (t,x), that every other player is certain of E at x, and so forth.

Define $CC(t,x) \equiv \cup_{m=0}^{\infty} T^m(t,x)$.

Let $A \subseteq T$. We say that there is CC of A at (t,x) (or, A is CC at (t,x)) if $CC(t,x) \subseteq A$.

Remark:

When the state t or the vertex x are obvious they might not be mentioned. So we might simply write "there is CC of A", or "t_i is certain of A". Usually we will refer to beliefs at the initial vertex. If the vertex is not mentioned it is the initial vertex.

The example below demonstrates how the informal discussion in the introduction can be mapped into the formal model.

Example 1

Consider the game that is described in figure 1. Table i,i = 1,2, describes the set of possible types, T_i, of player i. The left column describes the strategy of the given type and the other columns describe the beliefs on the type of the other player. So, for example, type t_1^1 of player I believes at x_0 that with probability 0.7 the type of player II is t_2^1 and with probability 0.3 his type is t_2^2. A complete description would include a specification of the beliefs in each vertex. But for our purposes it is sufficient to specify the beliefs of player I at x_0 and the beliefs of player II at x_0 and x_1.[7]

[7] Since player I moves at x_0 his beliefs at x_1 are the same as his beliefs in x_0.

Type of Player I	Strategy x_0	x_2	Beliefs at x_0
t_1^1	R	B	(0.7, 0.3)
t_1^2	B	B	(0.9,0.1)
t_1^3	R	R	(1,0)
t_1^4	R	B	(1,0)

Type of Player II	Strategy x_1	x_3	Beliefs x_0	x_1
t_2^1	B	B	(1,0,0,0)	(1,0,0,0)
t_2^2	R	B	(0,1,0,0)	(0,0,0.5,0.5))

It is easy to verify that types t_1^1 and t_1^2 for player I and types t_2^1 and t_2^2 for player II are rational. For example, the expected utility of type t_1^1 from playing the strategy (R,B) w.r.t. his beliefs is $(-1) \cdot 0.7 + 3 \cdot 0.3 = 0.2$. This is greater than his utility from playing B at the first stage. Consider the state $t = (t_1^1, t_2^1)$. We have $CC(t,x_0) = \{(t_1^1, t_2^1), (t_1^1, t_2^2), (t_1^2, t_2^1), (t_1^2, t_2^2)\}$. So there is CCR at (t, x_0). However, at x_1 type t_2^2 believes that player I is irrational and that with probability 0.5 he will play R and with probability 0.5 B. So it is possible to have CCR at the beginning of the game and not have it at a subsequent stage. Note that in this example type t_2^2 believes, at x_1, in theory 1, (page 4), and type t_2^1 is similar to the type that is described in theory 3.

3. COMMON CERTAINTY OF RATIONALITY

We have seen that CCR does not (necessarily) imply the backward induction outcome. In this section we show that the set of outcomes that are consistent with CCR at the beginning of the game coincides with the set of outcomes that survives one deletion of weakly dominated strategies and then iterative deletion of strongly dominated strategies. The formal description is as follows.

Let G be a game. Define:

$CCR(G) \equiv \{z(s)$: There exists a world T and a state $t \in T$ such that s is played in t and there is CCR at $(t,x_0)\}$;

CCR(G) is the set of outcomes that are consistent with CCR at the beginning of the game.

A strategy $\bar{s}_i \in S_i$ is <u>strongly dominated</u> if for every $\tau \in \Delta(S_{-i})$ $\bar{s}_i \notin \underset{s_i \in S_i}{\operatorname{argmax}}\, u_i(s_i,\tau)$.

A strategy $\bar{s}_i \in S_i$ is <u>weakly dominated</u> if for every $\tau \in \Delta(S_{-i})$ with full support (i.e. $\forall\, s_{-i} \in S_{-i}$ $\tau(s_{-i}) > 0$) $\bar{s}_i \notin \underset{s_i \in S_i}{\operatorname{argmax}}\, u_i(s_i,\tau)$.

We let $WS^{\infty}(G)$ denote the set of profiles of strategies $s \in S$ that survive one round of deletion of weakly dominated strategies and then iterative deletion of strongly dominated strategies. Formally, define:

$$WS_i^0(G) = \{s_i \mid s_i \in S_i,\, s_i \text{ is not weakly dominated in } G\}$$

For $n \geq 1$

$WS_i^n(G) = \{s_i \mid s_i \in WS_i^{n-1}(G)$ s_i is not strongly dominated in the subgame of G where each player j is restricted to strategies in $WS_j^{n-1}(G)\}$.

$$WS_i^{\infty}(G) \equiv \cap_{n=0}^{\infty} WS_i^n(G) .$$

$$WS^{\infty}(G) \equiv \Pi_{i=1}^{n} WS_i^{\infty}(G) .$$

We let $D(G)$ denote the set of outcomes that can be obtained by profiles of strategies in $WS^{\infty}(G)$. So $D(G) \equiv \{z(s) \mid s \in WS^{\infty}(G)\}$.

<u>THEOREM 1</u>: $CCR(G) = D(G)$.

The first part in the proof of the theorem shows that the set of outcomes that are consistent with the players' rationality coincides with the set of outcomes that survive one deletion of weakly dominated strategies. This part of the proof is established in Lemmas 1.1 and 1.2 below.

<u>DEFINITION</u>: Let $s_i, s_i' \in S_i$. The strategy s_i' is <u>equivalent</u> to the strategy s_i if for every $s_{-i} \in S_{-i}$ $z(s_i, s_{-i}) = z(s_i', s_{-i})$.

<u>LEMMA 1.1</u>: If $s_i \in S_i$ is not weakly dominated then there exists $s_i' \in S_i$ such that s_i' is equivalent to s_i and s_i' is rational.

<u>LEMMA 1.2</u>: If s_i is rational, then s_i is not weakly dominated.

The proofs of the lemmas are given in the appendix.[8]

<u>PROOF OF THEOREM 1</u>: We will show that $CCR(G) \subseteq D(G)$ and that $D(G) \subseteq CCR(G)$. Assume, first, that $z \in CCR(G)$. There exists a world T and a type $\bar{t} \in T$ such that $z = z(s(\bar{t}))$ and there is CCR at $(\bar{t}, x_0)$. We will show by induction that if $t_i \in [CC(\bar{t},x_0)]_i$ then $s_i(t_i) \in WS_i^k(G)$ for every $k = 0,1,\ldots$. For $K = 0$ the claim follows from lemma 1.2 and because t_i is rational. Assume now that for every $j \in N$ and every $t_j \in [CC(\bar{t},x_0)]_j$, $s_j(t_j) \in WS_j^k(G)$. It follows from the definition of $CC(\bar{t},x_0)$ that if $t_i \in [CC(\bar{t},x_0)]_i$ then t_i is certain that the type of player j, $j = 1,\ldots,n$, belongs to $[CC(\bar{t},x_0)]_j$. The induction hypothesis implies that t_i is certain that every player j is playing a strategy in $WS_j^k(G)$. It follows from the definition of $WS_i^{k+1}(G)$ and the rationality of t_i that $s_i(t_i) \in WS_i^{k+1}(G)$. We conclude that for every $t \in CC(\bar{t},x_0)$, $s(t) \in WS^\infty(G)$. In particular, $s(\bar{t}) \in WS^\infty(G)$ and therefore $z(\bar{t}) \in D(G)$.

We have shown that $CCR(G) \subseteq D(G)$. We now prove that $D(G) \subseteq CCR(G)$. Let $s_i \in WS_i^\infty(G)$ such that s_i is rational. There exists a probability distribution $\mu_i^1 \in \Delta(WS_{-i}^\infty(G))$ such that s_i is a best response to μ_i^1. It follows from lemma 1.1 that we can choose μ_i^1 so that its support contains only profiles s_{-i} which are rational (i.e. every strategy in s_{-i} is rational). Since s_i is rational there exists also a system of beliefs $\mu_i^2: X \to \Delta(T_{-i})$ such that for every $x \in X$ s_i is a best response to $\mu_i^2(x)$ in the game $G(x)$. We associate with s_i a type $t_i(s_i) = (s_i, \mu_i)$ where $\mu_i : X \to \Delta(T_{-i})$ is defined so that the following is satisfied: If x is consistent with some s_{-i} in the support of μ_i^1 then $\mu_i(x)[t_1(s_1),\ldots,t_{i-1}(s_{i-1}), t_{i+1}(s_{i+1}),\ldots,t_n(s_n)]$ is the probability distribution in $\Delta(T_{-i})$

[8] After a first draft of the paper was completed I learned that Battigalli (1990) has a different proof of lemma 1.2.

that is associated (via the correspondence $s_i <\!\!-\!\!> t_i(s_i)$), with the conditional probability of μ_i^1 given that the play of the game has reached the vertex x (the probability of other profiles of types is zero). Otherwise $\mu_i(x) = \mu_i^2(x)$. It is easy to check that $t_i(s_i)$ is a rational type. Now, let $z \in D(G)$. It follows from Lemma 1 that there exists a profile $s = (s_1,\ldots,s_n)$ such that $s \in WS^\infty(G)$, s_i is rational for every i, and $z = z(s)$. Consider the state $\bar{t} = (t_1(s_1),\ldots,t_n(s_n))$. It is easy to verify that there is CCR at $(\bar{t},x_0)$ and, of course, the outcome at t is z.

The two examples below demonstrate the theorem.

Example 2: Figure 2.2. describes the normal–form of the game that is described in figure 2.1. In this game the strategy ℓ is weakly dominated, and in the subgame that is left after the deletion of ℓ R is strongly dominated. So by theorem 1 CCR(G) = $\{z_2\}$.

It is easy to show that in a general two–stage game the only outcome that is consistent with CCR is the backward induction outcome. This follows because rationality implies that a player who moves in the second stage will choose the backward induction strategy, and since the player who moves in the first stage believes at the beginning that all the players are rational.

Example 3: Figure 3.2 describes the (reduced) normal–form of the game that is described in figure 3.1. In this game iterative deletion ends after the first stage, the only strategy that is deleted is RR. Thus, by theorem 1 CCR(G) = $\{z_1,z_2,z_3\}$. The outcome z_3 belongs to CCR(G) because at x_1 player II may believe that player I is not rational and will play R at x_2, while in fact, player I is rational and plays B at x_2. $z_2 \in$ CCR(G) because player I may believe that player II has the beliefs that were described above, while in fact, player II is not fooled by the move of player I.

4. NASH EQUILIBRIUM

In this section we show that players will play a Nash equilibrium path at a state $\bar{t} \in T$ if there is CC at $(\bar{t},x_0)$ of the following three events:

1. All the players are rational.
2. Each player gives a positive probability to the true state of the world.
3. The support of the beliefs of each player is the support of his beliefs at the actual state $\bar{t}$.

Formally, we define these events as follows:

$A_1 \equiv \{t \mid \forall i \in N \;\; t_i \text{ is rational}\}$

$A_2 \equiv \{t \mid \forall i \in N, \; u_i(t{-}i \mid t_i, x_0) > 0\}$

$A_3(\bar{t}) \equiv \{t \mid \forall i \in N \;\; \mu_i(t'_{-i} \mid t_i,x_0) > 0 \;\text{ iff } u_i(t'_{-i} \mid \bar{t}_i, x_0) > 0\}.$

<u>THEOREM 2</u>: If there is CC of A_1, A_2, and $A_3(\bar{t})$ at $(\bar{t},x_0)$ then:

1. The path $p(s(\bar{t}))$ is CC at $(\bar{t},x_0)$.
2. The path $p(s(\bar{t}))$ is a Nash equilibrium path.

The proof of the theorem is given at the end of the section. First, we discuss briefly the interpretation and the relative strength of our assumptions. Then we show, by means of examples, that these assumptions are indeed needed to obtain the result.

The simplest characterization of Nash equilibrium is as follows: If players are rational and if each one of them knows the strategies that will be played by the others, then the profile of strategies constitutes a Nash equilibrium. Roughly[9] speaking, our characterization substitutes the assumption that a player knows the strategies of the others, by the assumption that a player gives a positive probability to the profile of strategies of the others, and that a player knows the support of the beliefs of any other player. In addition, we assume that the content of these assumptions is CC.

A natural way of interpreting our assumptions is as follows: Assume that a player does not give a zero probability to an event unless there is compelling evidence that the event did not occur. The assumption that a player gives a positive probability to the true state means that a player will not receive compelling evidence that a state did not occur when it actually did. The assumption that a player knows the support of the beliefs of another player means that the former player observes any compelling evidence that the latter receives. (The emphasis is on compelling. A player might receive many other signals which affect his belief without changing its support. We do not assume that such signals are observed). We note that in a generic normal–form game the set of outcomes that are consistent with CCR is equal to the set of outcomes that are consistent with CCR, and CC of A_2 and $A_3(\bar{t})$ (where $\bar{t}$ is the actual state). So for a typical normal–form game, the assumptions that are made in theorem 2, do not restrict the outcome of the game to a set that is smaller than the set of outcomes that are consistent with CCR.

---

[9] "Roughly", because our assumption refers not to strategies but to types.

We now show that we cannot drop anyone of our assumptions and still obtain the result. The tables below refer to the game in figure 1.[10]

Type of Player I	Strategy x_0	Strategy x_2	Beliefs at x_0
t_1^1	R	B	(0.1, 0.9)
t_1^2	B	B	(0.9, 0.1)
t_1^3	R	R	–

Type of Player II	Strategy x_1	Strategy x_3	Beliefs x_0	Beliefs x_1
t_2^1	B	B	(0,1,0)	(1,0,0)
t_2^2	R	B	(0,1,0)	(0,0,1)

Consider the state $t = (t_1^1, t_2^1)$. $CC(t,x_0) = \{(t_1^1,t_2^1)(t_1^1,t_2^2), (t_1^2,t_2^1), (t_1^2, t_2^2)\}$. So there is CC of A_1 and $A_3(t)$, at x_0 yet player I plays R. Note that A_2 is not CC at (t,x_0) ($t \notin A_2$ because t_2^1 gives a probability zero to the true type of player I). It is not difficult to show that in this game player I will play B at the first move whenever the following two conditions are satisfied:

[10] The beliefs of type t_1^3 are irrelevant.

1. There is CCR at x_0.
2. Player I gives (at x_0) a probability 1 to the event that player II gives (at x_0) a positive probability to the actual move of player I.

Note that the second condition is implied by CC of A_2. It can be shown that in a general centepede game (i.e. a centepede game of any length) CCR and CC of A_2 imply that player I plays B at the first move. (B in the first move is the only Nash equilibrium path in this game). The next example shows that, in general, CCR and CC of A_2 do not imply a play along a Nash equilibrium path. The tables below refer to the game that is described in figure 4.[11]

Type of Player I	Strategy			Beliefs		
	x_0	x_2	x_3	x_0	x_2	x_3
t_1^1	R	L	L	(0.5,0,0.5,0,0)	(1,0,0,0,0)	(0,0,0,1,0)
t_1^2	R	R	R	(0,0.5,0.5,0,0)	(0,0,0,0,1)	(0,1,0,0,0)

[11] Only the relevant beliefs have been specified.

Type of Player II	Strategy x_0	x_4	x_5	Beliefs x_0 and x_1	x_4	x_5
t_2^1	L	L	R	(1,0)	(1,0)	(0,1)
t_2^2	M	L	R	(0,1)	(1,0)	(0,1)
t_2^3	R	L	R	(0.5,0.5)	(1,0)	(0,1)
t_2^4	M	R	L	–	–	–
t_2^5	L	R	L	–	–	–

Consider the state $t = (t_1^1, t_2^3)$.

It is easy to see that $CC(t,x_0) = \{(t_1^1,t_2^1), (t_1^1,t_2^3), (t_1^2,t_2^2), (t_1^2,t_2^3)\}$ and therefore $CC(t,x_0) \subsetneq A_1 \cap A_2$. Thus, at (t, x_0) there is CC of A_1 and A_2. However, $p(s(t))$ is not a Nash equilibrium path (player I would benefit from deviating and playing L at the beginning). Note that at (t, x_0), player II does not know the support of the beliefs of player I, player II believes that with probability 0.5 player I will be surprised by the move L and with probability 0.5 he will be surprised by the move M.

Before presenting the formal proof of theorem 2 we give a sketch of the main steps in the argument. Lemma 2.1 provides a simple characterization of Nash equilibrium paths. This characterization implies that if each player is rational and certain of the true path, then the path is a Nash equilibrium path. Lemma 2.2, (which is the main lemma), shows that if the provisions in the theorem are satisfied, then the set $CC(\bar{t},x_0)$ has a certain simple structure. This structure implies that the path $p(s(\bar{t}))$ is CC at $(\bar{t},x_0)$.

Specifically, this structure implies the following property: Let $x_1, y_1, \ldots, y_m \in X$ be vertices that are consistent with $CC(\bar{t}, x_0)$ such that $y_1, \ldots, y_m$ are successors of x. Let $i(x)$ denote the player that moves at x. If for every $j = 1, \ldots, m$, there is CC of the path that will be played given that y_j has been reached, then there is CC that at x that $i(x)$ will move to the vertex y_k, $k \in \{1, \ldots, m\}$, that corresponds to the path that maximizes the utility of player i. It follows that $m = 1$, i.e., there is only one successor of x that is consistent with $CC(\bar{t}, x_0)$. So for $t \in CC(\bar{t}, x_0)$ such that x is consistent with t $p(s(t))$ is CC at (t,x). A simple induction on the height of the vertices implies that $p(s(\bar{t}))$ is CC at $(\bar{t}, x_0)$.

We now turn to the formal proof:

<u>LEMMA 2.1</u>: A path $\bar{p}$ is a Nash equilibrium path iff for every $i \in N$ and every $x \in X_i$ on the path $v_i(x) \leq u_i(z(\bar{p}))$.

<u>PROOF</u>: Let $s = (s_1, \ldots, s_n)$ be a Nash equilibrium such that $p(s) = \bar{p}$. Assume by contradiction that there exists a vertex $x \in X_i$ on the path $\bar{p}$ such that $v_i(x) > u_i(z(\bar{p}))$. Player i will improve his payoff if he deviates to a strategy s_i' which plays like s_i until x is reached and then plays the maxmin strategy in $G(x)$. This contradicts our assumption that s is a Nash equilibrium. Assume now that $v_i(x) \leq u_i(z(\bar{p}))$ for every i and every $x \in X_i$ on the path. Define $s = (s_1, \ldots, s_n)$ so that players follow the path $\bar{p}$ as long as possible. If a player deviates from $\bar{p}$ then the others minmax him. Since the maxmin, $v_i(\ )$, equals the minmax, it is easy to see that s is a Nash equilibrium and $p(s) = \bar{p}$.

LEMMA 2.2: Let $\bar{t} \in T$. If A_2 and $A_3(\bar{t})$ are CC at $(\bar{t}, x_0)$ then there exists sets $\bar{T}_i \subseteq T_i$, $i = 1,\ldots,n$, such that

$$CC(\bar{t}, x_0) = \Pi_{i=1}^{n} \bar{T}_i$$

and for every $t_i \in \bar{T}_i$ and $t_{-i} \in \Pi_{j\neq i} \bar{T}_j$ $\mu_i(t_{-i} \mid t_i,x_0) > 0$.

The proof of the lemma is based on the following lemmatas, which assume the provisions of the lemma.

Define $\bar{T}_k^i \equiv \{t_k \mid t_k \in T_k,\ \mu_i(t_k \mid \bar{t}_i,x_0) > 0\}$.

LEMMATA 2.2.1: For every $i,j,k \in \{1,\ldots,n\}$, $k \neq i,j$, $\bar{T}_k^i = \bar{T}_k^j$.

PROOF: We will show that $\bar{T}_k^i \subseteq \bar{T}_k^j$. Let $t_k^{1'} \in \bar{T}_k^i$, there exists $t \in CC(\bar{t},x_0)$ such that $t_k = t_k'$, $t_i = \bar{t}_i$ and $\mu_i(t_{-i}|\bar{t}_i,x_0) > 0$. Since there is CC of A_2 $\mu_j(t_{-j}|t_j,x_0) > 0$. Since $A_3(\bar{t})$ is CC $\mu_j(t_{-j}|\bar{t}_j,x_0) > 0$. This means that $t_k \in \bar{T}_k^j$. In a similar way we can show that $\bar{T}_k^j \subseteq \bar{T}_k^i$. It follows that $\bar{T}_k^i = \bar{T}_k^j$.

LEMMATA 2.2.2: For every $t_{-i} \in \Pi_{j\neq i} \bar{T}_j^i$, $i = 1,\ldots,n$, $\mu_i(t_{-i}|\bar{t}_i,x_0) > 0$.

PROOF: Assume w.l.o.g. that $i = 1$. Let $\tilde{t}_{-1} \in \Pi_{j=2}^{n} \bar{T}_j^1$. Define $A_k \equiv \{t|\ t_1 = \bar{t}_1$, and $t_\ell = \tilde{t}_{-1,\ell}$, $2 \leq \ell \leq k\}$. Let m be the maximal number k such that player 1 gives a positive probability to the event A_k, at $(\bar{t}, x_0)$. We wish to show that $m = n$. Assume by contradiction that $m < n$. There exists $t \in A_m$ such that $\mu_1(t_{-1} | \bar{t}_1, x_0) > 0$. Let $h = m+1$. Since A_2 is CC. $\mu_h(t_{-h} | t_h, x_0) > 0$. Since $A_3(\bar{t})$ is CC the beliefs of t_h and $\tilde{t}_h$ have the same support, therefore $\mu_h(t_{-h} | \tilde{t}_h, x_0) > 0$. Now again, because A_2 is CC, $\mu_1((\tilde{t}_2, \ldots, \tilde{t}_h, t_{h+1}, \ldots, t_n) \mid \bar{t}_1, x_0) > 0$. So we have obtained a contradiction to the assumption that $m < n$. The lemmata follows.

PROOF of the LEMMA: Define $\bar{T}_i \equiv \bar{T}_i^1$ for $i \neq 1$ and $\bar{T}_1 = \bar{T}_1^2$. The lemmetas imply that for every $i \in N$, and every $t_{-i} \in \Pi_{j \neq i} \bar{T}_j$ $\mu_i(t_{-i} | \bar{t}_i, x_0) > 0$. Conversely, if $\mu_i(t_{-i} | \bar{t}_i, x_0) > 0$ then $t_{-i} \in \Pi_{j \neq i} \bar{T}_j$. Let $t_i \in T_i$. Define $B(t_i) \equiv \{t_{-i} | \mu_i(t_{-i} | t_i, x_0) > 0\}$. Since $A_3(\bar{t})$ is CC $B(t_i) = \Pi_{j \neq i} \bar{T}_j$ for every $t_i \in \bar{T}_i$. A simple induction shows that $T^m(\bar{t}, x_0) = \pi_{i=1}^n \bar{T}_i$ for $m \geq 2$. It follows that $CC(\bar{t}, x_0) = \Pi_{i=1}^n \bar{T}_i$.

PROOF OF THE THEOREM: First we note that if the players are rational and if each player is certain at $(\bar{t}, x_0)$ that the outcome of the game will be the path $p(s(\bar{t}))$, then the condition in lemma 2.1 is satisfied. This follows because:

1. Along the path $p(s(\bar{t}))$ players maintain their initial belief.
2. At a vertex $x \in X_i$ player i can guarantee a payoff of $v_i(x)$ (and therefore will not follow a path which gives him less than that).

We conclude that the second part of the theorem follows from the first part. We now prove the first part. By lemma 2.2 there exists sets $\bar{T}_i$, $\bar{T}_i \subseteq T_i$,

$i = 1,\dots,n$, such that $CC(\bar{t},x_0) = \Pi_{i=1}^{n}\bar{T}_i$ and for every $t_i \in \bar{T}_i$ and $t_{-i} \in \bar{T}_{-i}$ $\mu_i(t_{-i}|t_i,x_0) > 0$. It follows that for every $t \in CC(\bar{t},x_0)$ and $x \in p(s(t))$ $CC(t,x) = \Pi_{i=1}^{n}\bar{T}_i(x)$ where

$$\bar{T}_i(x) \equiv \{t_i | t_i \in \bar{T}_i \text{ and } t_i \text{ is consistent with } x\}.$$

Let $t \in CC(\bar{t},x_0)$ and let $p(s(t) = x_0(t), x_1(t),\dots,x(t)_{m(t)}, (x_0(t) = x_0)$. Define:

$$h(t) \equiv \min\{k| \text{ there is CC at } (t,x_k(t)) \text{ that the path is } p(s(t))\}.$$

$$q(\bar{t}) \equiv \max\{h(t)|\ t \in CC(\bar{t},x_0)\}.$$

We want to show that $h(\bar{t}) = 0$. Assume by contradiction that $h(\bar{t}) > 0$, it follows that $q(\bar{t}) > 0$. Define $b \equiv q(\bar{t}) - 1$. Let $t \in CC(\bar{t},x_0)$ be a state such that at $(t,x_b(t))$ there is no CC of the path $p(s(t))$. Assume that $x_b(t) = x_i$ where $x_i \in X_i$. Let x be a successor of x_i that is consistent with some $t' \in CC(\bar{t},x_0)$. It follows from the choice of x_i that we can associate with x a path p_x such that p_x is CC at (t',x). Since $CC(t',x) = \Pi_{j=1}^{n}\bar{T}_j(x)$ it follows that every $t_j \in \bar{T}_j(x)$, $j = 1,\dots,n$, is consistent with p_x. Since $\bar{T}_j(x) = \bar{T}_j(x_i)$ for every $j \neq i$ there is CC at (t,x_i) that player i can play in a way that will lead to the outcome $z(p_x)$. Define:

$$B \equiv \left\{ x \,\middle|\, \begin{array}{l} x \text{ is a successor of } x_i \text{ and} \\ x \text{ is consistent with } CC(\bar{t},x_0) \end{array} \right\}.$$

Since there is CCR at (t,x_i) it follows that there is CC at (t,x_i) that $s_i(x_i) = \underset{x \in B}{\text{argmax}}\, \mu_i(z(p_x))$. This implies that B has a single element. It follows that $p(s(t))$ is CC at (t,x_i). Thus we have obtained a contradiction to the assumption that $q(\bar{t}) > 0$. It follows that $p(s(\bar{t}))$ is CC at $(\bar{t},x_0)$.

The assumptions in theorem 2 do not, in general, imply the backward induction outcome. To see this, consider the game G that is described in figure 5.[12] The backward induction path is (x_0,x_1,x_3,z_1). However, the path (x_0,x_2,x_4,z_2), which is a Nash equilibrium path can be the outcome of the game in a state t which satisfies the assumptions in theorem 2. Specifically, at (t,x_0) player II is certain that player I will play to x_2. If player I plays to x_1, player II believes that I is irrational and that at x_3 I will play R, therefore at x_1 II plays R. The full description is given in the following tables.

Type of Player I	Strategy x_0	x_3	x_4	Beliefs
t_1^1	R	L	R	1
t_1^2	L	R	L	–

[12] Reny (1988) considers a similar example.

Type of Player II	Strategy x_1	x_2	x_0	Beliefs x_1	x_2
t_2^1	R	R	(1,0)	(0,1)	(1,0)

At the state $t = (t_1^1, t_2^1)$ the path (x_0, x_2, x_4, z_2) is played, and it is easy to verify that there is CC of A_1, A_2 and $A_3(t)$.

As of yet I was unable to find an interesting characterization of backward induction in a model without mistakes. In the next section we introduce the possibility of mistakes in the implementation of the strategies and show how this possibility can select the backward induction outcome.

5. BACKWARD INDUCTION

In this section we extend the model and introduce the possibility of mistakes in the implementation of the strategies. It is shown that if, at the beginning of the game there is CCR and CC that there is a "small" probability for a mistake, then the players choose the backward induction strategies. The reason for this is simple: A move to a vertex which is not consistent with CCR, in a world without mistakes, can now be attributed to a mistake in the implementation of the strategy, and therefore if there is CCR at the beginning there is CCR at every vertex and then the backward induction argument is valid.

We now turn to a formal description. A state of the world should specify the mistakes that occur. So a state of the world specifies a function φ: S x X $\rightarrow$ X, where $\varphi(s,x)$ is a successor of x. The function φ determines what will occur at the vertex x when the profile that is played is s. We assume that at a given state the move that will occur at a vertex x depends only on the move that player $i(x)$ planned to play.[13] Formally, $\varphi(s,x) = \varphi(s',x)$ whenever $s_{i(x)} = s'_{i(x)}(x)$. Thus, the set of states of the world is $\Omega = T \times \phi$, where $T = \pi_{i=1}^{n} T_i$ is the set of profiles of types of the players and ϕ is a set of functions from $S \times X$ to X such that each function satisfies the property that has been specified above. A type $t_i \in T_i$ for player i specifies a pair (s_i, μ_i) where $s_i \in S_i$ is a strategy for player i and $\mu_i: X \to \Delta(T_{-i} \times \phi)$ is a system of beliefs. A profile of strategies s and a function φ determin an outcome $z \in Z$ in an obvious way. We let $\varphi(s)$ denote this outcome. The definitions of rationality and CC are obvious modifications of the definition in section 2. So, for example, the first requirement in the definition of rationality is that at a vertex x the player gives a positive probability only to states (t,φ) in which x can be reached. To define CC of an event $A \subseteq T \times \phi$ at a state $w = (t,\varphi)$ at a vertex x we define $\Omega^m(w,x)$ in a way that is similar to the definition of $T^m(t,x)$. So

$$\Omega^0(w,x) \equiv \{w\},$$

$$\Omega^m(w,x) \equiv \{(t',\varphi') \mid \text{for some } i \in N \;\; t'_i \in \Omega_i^{m-1}(w,x) \text{ and } \mu_i((t'_{-i},\varphi') \mid t'_i,x) > 0\},$$

$$CC(w,x) \equiv \bigcup_{m=0}^{\infty} \Omega^m(w,x),$$

and then A is CC at (w,x) if $CC(w,x) \subseteq A$.

We now define the events that are related to our assumptions on beliefs about mistakes. Let $\mu_j(A \mid t_j,x)$ denote the probability that type t_j gives at the vertex x to the event $A \subseteq T \times \phi$. Define:

$$A_x \equiv \{(t,\varphi) \mid \varphi((s'_i, s_{-i}(t_{-i})), x) = s'_i(x) \text{ for } s'_i \in S_i \text{ where } i = i(x)\}$$

$$D_z \equiv \{(t,\varphi) \mid \varphi(s(t)) = z)$$

---

[13] As will become clear in the sequel the model does allow beliefs in which a specific mistake is correlated with the whole profile of strategies and not only with a move at the corresponding vertex.

$B(\epsilon) \equiv \{(t,\varphi) \mid \mu_j(A_y \mid t_j, x) \geq 1-\epsilon$ for every $j \in N$, $x \in X$, and y that is a vertex in the game $G(x)\}$

$D \equiv \{t,\varphi) \mid \mu_j(D_z \mid t_j,x_0) > 0$ for every $j \in N$ and $z \in Z\}$.

D is the event where, at the beginning of the game, each player gives a positive probability for every outcome (such a belief will follow, for example, from a belief that every mistake can occur with a positive probability.) $B(\epsilon)$ is the event where for every vertex x, each player believes that the probability for a mistake at any given vertex in the sequel is at most ϵ.

<u>THEOREM 3</u>: There exists an $\bar{\epsilon}$ such that for every $0 < \epsilon < \bar{\epsilon}$ the following is satisfied: If D, $B(\epsilon)$ are CC at $((\bar{t},\bar{\varphi}),x_0)$ and if there is CCR at $((\bar{t},\bar{\varphi}),x_0)$ then $s(\bar{t})$ is the profile of backward induction strategies.

Thus, to obtain the backward induction outcome we (only) have to assume CCR and CC that every player believes that every mistake is possible, but occurs with a small probability. Our assumptions allow (while Selten (1975) did not), the belief of player i about a mistake by player j to be correlated with his beliefs about the profile of strategies that is played. In particular, the belief of player i can correlate between the mistakes of player j. Therefore player i can believe, for example, that if player j has made one mistake he is more likely to make another one.

The proof of the theorem requires the following lemma.

<u>LEMMA 3.1</u>: Let $w = (t,\varphi)$. If D is CC at (w,x_0) and A is some other event that is CC at (w,x_0) then A is CC at (w,x) for every x.

<u>PROOF</u>: We will prove that $CC(w,x) \subseteq CC(w,x_0)$ by showing that for every m $\Omega^m(w,x) \subseteq \Omega^m(w,x_0)$. We have $\Omega^0(w,x) = \Omega^0(w,x_0) = w$. Now, assume by induction that the claim is true for m–1 and let $(t',\varphi') \in \Omega^m(w,x)$. There exists a player i such that $\mu_i((t'_{-i},\varphi) \mid t'_i,x) > 0$ and $t'_i \in \Omega^{m-1}(w,x)$. By the induction hypothesis $t'_i \in \Omega^{m-1}(w,x_0)$ and since D is CC at (w,x_0) t'_i gives a positive probability at x_0 to the event where the vertex x is reached. Since t'_i updates his beliefs according to Bayes' law we have $\{(t_{-i},\varphi) \mid \mu_i((t_{-i},\varphi)|t'_i, x) > 0\} \subseteq \{(t_{-i},\varphi) \mid \mu_i((t_{-i},\varphi)|t'_i, x_0\} > 0\}$ and therefore $(t',\varphi) \in \Omega^m(w,x_0)$.

□

<u>Remark</u>: Let $w' \in CC(w,x_0)$. It is easy to see that $CC(w',x_0) \subseteq CC(w,x_0)$. It follows that the provisions of the lemma are satisfied at w' as well and therefore $CC(w',x) \subseteq CC(w',x_0)$ for every x.

We now turn to the proof of the theorem. Let:

$$R \equiv \max_{i \in N} \max_{z,w \in Z} | u_i(z) - u_i(w) |$$

$$r \equiv \min_{i \in N} \min_{\substack{z,w \in Z \\ z \neq w}} | u_i(z) - u_i(w) | .$$

Let k denote the height of the game. Define $\bar{\epsilon}$ to be the number which satisfies the following equation $[1-(1-\bar{\epsilon})^k] \cdot R = (1-\bar{\epsilon})^k r$. We will show by induction on h(x), (the height of the vertex x), that for every x and every $(t,\varphi) \in CC((\bar{t},\bar{\varphi}),x_0)), s_i(t_i)$

prescribes the backward induction move at x. So let $(t,\varphi) \in CC((\bar{t},\bar{\varphi}),x_0)$, let x be a vertex of height m and assume that the claim is true for vertices of a lower height. Let $i = i(x)$. Since D is CC at (t,φ) it follows from lemma 3.1 and from the induction hypothesis that at x t_i gives a probability 1 to the event where the other players play the backward induction strategies in $G(x)$. Since $(t,\varphi) \in B(\epsilon)$, t_i believes at x that the probability for a mistake at any vector y in $G(x)$ is at most ϵ. It follows that if i plays the backward induction strategy at $G(x)$ then his expected payoff (w.r.t. his beliefs), is bounded below by $(1-\epsilon)^m \cdot u_i(z) + [1 - (1-\epsilon)^m] \cdot r_i(x)$ where z is the backward induction outcome in $G(x)$ and where $r_i(x)$ is the lowest payoff for player i in the subgame $G(x)$. Similarly if i plays to a vertex x' which is not consistent with backward induction, then his payoff is bounded from above by $(1-\epsilon)^m \cdot u_i(z') + [1 - (1-\epsilon)^m] R_i(x)$ where z' is the backward induction outcome in $G(x')$ and $R_i(x)$ is the maximal payoff for player i in $G(x)$. Since $u_i(z') < u_i(z) + r$, $\epsilon < \bar{\epsilon}$ and $m < k$ it follows that the expected payoff of type t_i is maximized if he plays the backward induction move.

APPENDIX

PROOF OF LEMMA 1.1: Since s_i is not weakly dominated there exists a belief $\mu_i \in \Delta(S_{-i})$ with full support such that s_i is a best response to μ_i. Since μ_i has full support, the conditional probability of μ_i is defined for every vertex $x \in X$. Let μ_x denote this conditional probability. Let $x \in X_i$. Define s_i' as follows. If $x \in X(s_i)$, then $s_i'(s) \equiv s_i(x)$. If $x \notin X(s_i)$ then $s_i'(x)$ is a successor of x that is consistent with a best response to μ_x in $G(x)$. It is easy to see that s_i can be defined in such a way. (Define s_i' inductively from the end of the game tree to the initial vertex). Define $\mu_i'(x) \equiv \mu_x$. It is easy to see that s_i' is a best response to μ_i', therefore it is rational. Since s_i' differs from s_i only on vertices which are inconsistent with s_i, s_i' is equivalent to s_i.

The proof of lemma 1.2 relies on the following lemmata.

LEMMATA 1.2.1: Let s_i be a rational strategy for player i. For every $x \in X$ there exists a profile $\bar{s}_{-i} \in S_{-i}(x)$ such that s_i is a best response to $\bar{s}_{-i}$ in the game $G(x)$.

PROOF: Let $x \in X$. Since s_i is rational, there exists a belief $\mu_i(x)$ on $S_{-i}(x)$ such that s_i maximizes the expected utility of player i w.r.t. the belief $\mu_i(x)$. Since at x player i can guarantee himself $v_i(x)$, his expected utility (w.r.t. $\mu_i(x)$) is greater or equal to $v_i(x)$. Therefore, there is at least one profile, s_{-i}, in the support of $\mu_i(x)$ such that $U_i[z(s_i, s_{-i})] \geq v_i(x)$. It follows that there exists a path $p(x)$ from x to a terminal vertex z such that: 1. s_i is consistent with $p(x)$; 2. $v_i(x) \leq u_i(z)$. A simple induction on the height of x in the game tree shows the following: There exists a path $\bar{p}(x)$ from x to a terminal vertex $\bar{z}$ such that for every x' on $\bar{p}(x)$, $v_i(x') \leq u_i(\bar{z})$. To see this define y to be the highest vertex on the path $p(x)$ with the property that $u_i(z) < v_i(y)$. (If there is no such y $\bar{p}(x) = p(x)$.) By the induction hypothesis there exists a path $\bar{p}(y)$ from y to some terminal vertex w that satisfies $u_i(w) \geq v_i(y')$ for every y' on the path $\bar{p}(y)$. We set $\bar{z} \equiv w$, so that $\bar{p}(x)$ is the path from x to w. It is easy to see that $v_i(x') \leq u_i(\bar{z})$ for every x' on $\bar{p}(x)$. Define $\bar{s}_{-i}$ as follows i. $\bar{s}_{-i}$ is consistent with the path $\bar{p}(x)$. If player i deviates from $\bar{p}(x)$ then s_{-i} brings player i down to the maxmin, i.e. for every $y \in G(x)$ that is not on the path $\bar{p}(x)$, $\max_{s_i \in S_i(y)} u_i(s_i, \bar{s}_{-i}) = v_i(y)$.

We now show that s_i is a best response to s_{-i} in the game $G(x)$. Consider a deviation of player i from the path $\bar{p}(x)$. So suppose that at some $x_i \in X_i$ which is on the path $\bar{p}(x)$, player i deviates and plays to a vertex y which is not on the path. Since y is a successor of x_i $v_i(y) \leq v_i(x_i)$. By the construction of $\bar{p}(x)$ $v_i(x_i) \leq u_i(\bar{z})$. Thus, player i cannot gain from such a deviation.

□

<u>PROOF OF LEMMA 1.2</u>: We need to construct a belief $\mu \in \Delta(S_{-i})$ with full support, such that s_i is a best response to μ. Define:

$$\eta \equiv \min_{\substack{z, w \in Z \\ z \neq w}} | u_i(z) - u_i(w) |$$

$$w \equiv \max_{z, w \in Z} | u_i(z) - u_i(w) | .$$

In the class of games that we consider, $\eta > 0$ for every game. By Lemmeta 1.2.1. we can associate with each vertex $x \in X$ a profile $\bar{s}_{-i}(x) \in S_{-i}(x)$ such that for every $s_i' \in S_i(x)$ such that $z(s_i', \bar{s}_{-i}(x)) \neq z(s_i, \bar{s}_{-i}(x))$, $u_i(z(s_i, s_{-i}(x))) - \eta \geq u_i(z(s_i', \bar{s}_{-i}(x))$. We will construct μ so that for every $x \in X(s_i) \cap X_i$ the move $s_i(x)$ is consistent with a best response to μ_x in the game $G(x)$. (μ_x is the conditional of μ at x). It is easy to see that this implies that s_i is a best response to μ. Let $x \in X_i$. We claim that if μ_x satisfies the inequality (1) below, then $s_i(x)$ is consistent with a best response to μ_x in $G(x)$.

(1) $\quad (1 - \mu_x[\bar{s}_{-i}(x)])W < \mu_x[\bar{s}_{-i}(x)]\eta.$

To see this, compare the payoff of player i from a strategy s_i^* that is consistent with the path $p(s_i, \bar{s}_{-i}(x))$ with his payoff from a strategy s_i' such that $s_i'(x) \neq s_i(x)$. With probability $\mu_x[\bar{s}_{-i}(x)]$ the payoff of player i from s_i^* is greater than his payoff from s_1' by at least η, and with probability $1 - \mu_x[\bar{s}_{-i}(x)]$ his payoff from s_i^* will be lower than his payoff from s_i' by at most W. If (1) is satisfied, the payoff from s_i^* is higher. We will now construct μ so that for every $x \in X(s_i)$ μ_x satisfies the inequality (1). Define inductively a sequence of sets of vertices $X^n \subsetneq X(s_i)$ and a sequence of sets of strategies $S_{-i}^n \subsetneq S_{-i}$ as follows:

$X^1 \equiv \{x_0\}$

$S_{-i}^1 \equiv \{\bar{s}_{-i}(x_0)\}$

$X^n \equiv \{x | x \in X(S_i) - \bigcup_{j=1}^{n-1} X^j$ there exists a profile $s_{-i} \in S_{-i}^{n-1}$ and a vertex y such that y is on the path $p(s_i, s_{-i})$, $y \notin X_i$ and x is a successor of $y\}$

$S^n_{-i} \equiv \{\bar{s}_{-i}(x) \mid x \in X^n\}$.

The set X^n is the set of vertices that are obtained by a single deviation of a player $j \neq i$ from a path $p(s_i, s_{-i})$ where $s_{-i} \in S^{n-1}_{-i}$. The definition is complete at stage k where $X^{k+1} = \phi$. Let m be the number of vertices in the game tree and let $\epsilon > 0$ be a number which satisfies $m \cdot \epsilon \cdot W < (1-\epsilon)\cdot\eta$. We define μ as follows: The probability of the set S^n_{-i}, $n = 1,\ldots,k$, is $\epsilon^{n-1} - \epsilon^n$ and all the strategies in the set have the same probability. The probability of the set $S_{-i} - \cup^k_{n=1} S^n_{-i}$ is ϵ^k and again all the strategies in the set have the same probability. Now consider a vertex $x \in X(s_i)$. We want to show that the inequality in (1) is satisfied. It follows from the definitions of the sequences X^n and S^n_{-i} that there exists a number n and a vertex $x' \in X^n$ such that x is on the path $p(s_i, \bar{s}_{-i}(x'))$. The proof of lemmata 1.2.1. shows that we can assume that $\bar{s}_{-i}(x) = \bar{s}_{-i}(x')$. Since there are at most m strategies in S^n_{-i} (because there are at most m vertices in X^n), $\mu(\bar{s}_{-i}(x)] \geq \frac{1}{m}(1-\epsilon)\epsilon^{n-1}$. On the other hand $\mu\{S_{-i}(x) - \bar{s}_{-i}(x)] \leq \epsilon^n$. We have $\mu[\bar{s}_{-i}(x)]/\mu[S_{-i}(x) - \bar{s}_{-i}(x)] = \mu_x[\bar{s}_{-i}(x)] \,/\, 1 - \mu_x[\bar{s}_{-i}(x)]$. It follows from the choice of ϵ that $\mu_x[\bar{s}_{-i}(x)]$ satisfies (1).

REFERENCES

1. Aumann, R.: "Agreeing to Disagree", The Annals of Statistics 4, 1236–1239.
2. Aumann, R.: "Correlated Equilibrium as an Expression of Bayesian Rationality", Econometrica, 55 (1987) 1–18.
3. Aumman, R. and A.Brandenburger: "Epistemic Conditions for Nash Equilibrium." Working Paper No.91–042 (1991), Harvard Business School.
4. Battigalli, P.: "On Rationalizability in Extensive Games", mimeo, Universita Commerciale "L.Bocconi" (1990).
5. Bernheim, D.: "Rationalizable Strategic Behavior", Econometrica, 52 (1984), 1007–1028.
6. Bicchieri, C.: "Backwards Induction without Common Knowledge", Proceedings of the Philosophy of Science Association Meeting (1988).
7. Binmore, K.: "Modeling Rational Players, part I." Journal of Economics and Philosophy, 3 (1987), 179–214.
8. Borgers, T.: "Weak Dominance and Approximate Common Knowledge", mimeo Institut fur Volkswirtschaft, Universitat Basel (1990).
9. Borgers, T.: "On the Definition of Rationalizability in Extensive Games", Discussion Paper No.91–22, 1991, University College, London.
10. Brandenburger, A.: "Lexicographic Probabilities and Iterated Admissability", mimeo, Harvard Business School (1990).
11. Brandenburger, A.: "Knowledge and Equilibrium in Games", forthcoming in Journal of Economic Perspectives.

12. Brundenberger, A. and E. Dekel.: "Rationalizability and Correlated Equilibria", Econometrica, 55 (1987), 1391–1402.

13. Dekel, E. and D. Fudenberg.: "Rational Behavior with Payoff Uncertainty", J.E.T, 52 (1990), 243–267.

14. Gilboa, I.: "A Note on the Consistency of Game Theory", Discussion Paper No.847 (1989), MEDS/KGSM Northwestern University.

15. Gul, F.: "Rational Strategic Behavior and the Notion of Equilibrium", mimeo, GSB Stanford University (1989).

16. Kreps, D. and R.Wilson.: "Sequential Equilibria", Econometrica, 56 (1982), 863–894.

17. Monderer, D. and D.Samet.: "Approximating Common Knowledge with Common Beliefs", Games and Economic Behavior, 1 (1989), 170–190.

18. Pearce, D.: "Rationalizable Strategic Behavior and the Problem of Perfection", Econometrica, 52 (1984), 1008–1050.

19. Reny, P.: "Rationality Common Knowledge and the Theory of Games", mimeo, The University of Western Ontario (1988).

20. Rosenthal, R.: "Games of Perfect Information, Predatory Pricing and the Chain–Store Paradox", J.E.T., 25 (1981), 92–100.

21. Setlen, R.: "Re–examination of the Perfectness Concept for Equilibrium Points in Extensive Games", International Journal of Game Theory, 4, 1975, 25–55.

22. Selten, R.: "The Chain–Store Paradox," Theory and Decision, 9, (1978), 127–59.

23. Ten, T. and S.Werlang.: "The Bayesian Foundations of Solution Concepts of Games", J.E.T., 45 (1988), 370–391.

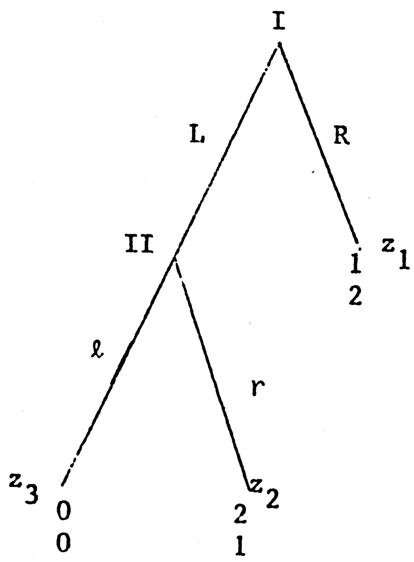

Figure 2.1

	ℓ	r
L	0,0	2,1
R	1,2	1,2

Figure 2.2

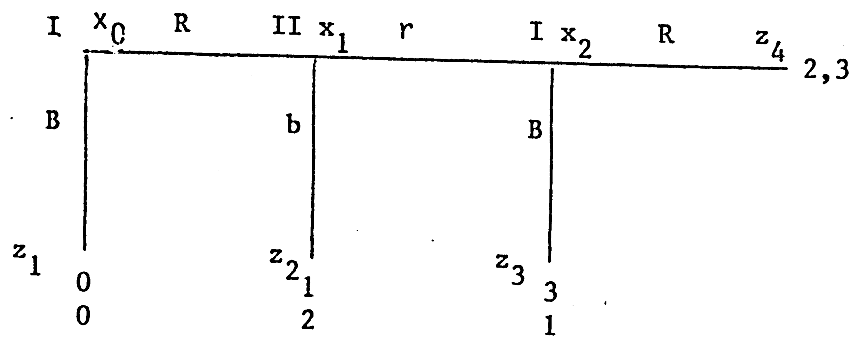

Figure 3.1

	b	r
BB	0,0	0,0
RB	-1,2	3,1
RR	-1,2	2,3

Figure 3.2

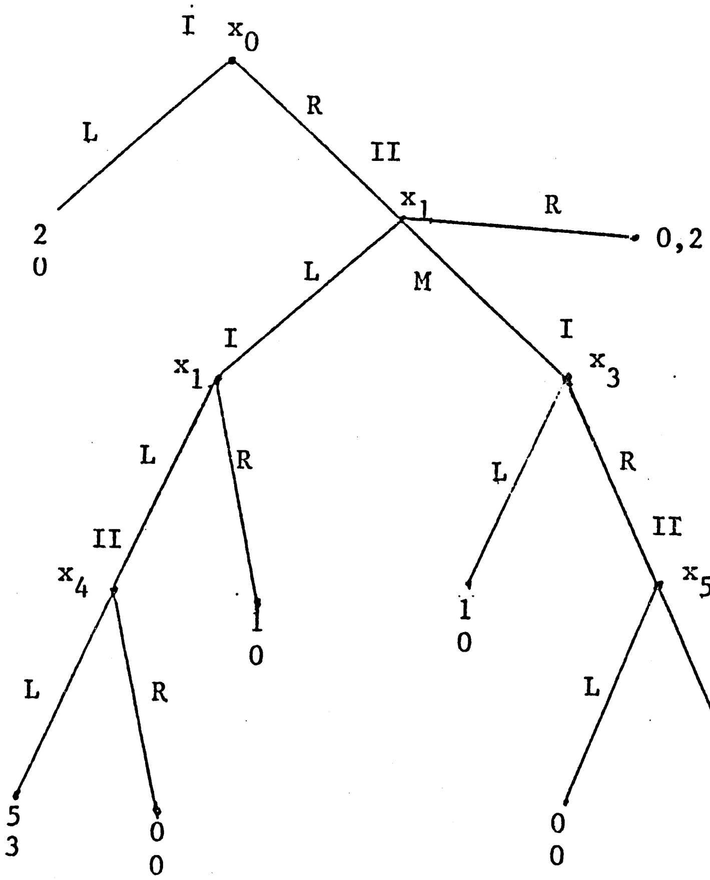

Figure 4

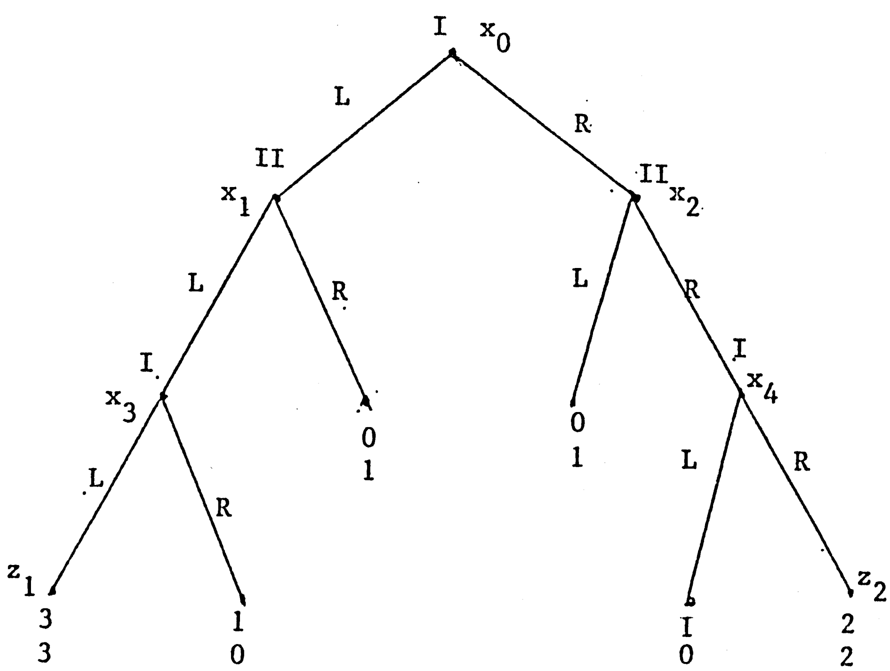

Figure 5

THE FOERDER INSTITUTE FOR ECONOMIC RESEARCH
List of Recent Working Papers*

29.91	Zvi Hercowitz L(Rubin)Meridor	The Macroeconomic Effects of Mass Immigration to Israel
30.91	Salvatore Modica	Neo–Bayesian Decision Making with Subjective Models
31.91	Gideon Fishelson	More on the Salant Principle
32.91	Chaim Fershtman Yoram Weiss	Social Status, Culture and Economic Performance
33.91	Chaim Fershtman Yoram Weiss	On the Stability of Occupational Rankings
34.91	Alex Cukierman Nissan Liviatan	The Dynamics of Optimal Gradual Stabilizations
35.91	Ariel Rubinstein	On Price Recognition and Computational Complexity in a Monopolistic Model
36.91	Neil Gandal David J.Salant	Movie Wars
37.91	Chaim Fershtman Neil Gandal	Disadvantageous Semicollusion
38.91	Zvi Eckstein Itzhak Zilcha	The Effects of Compulsory Schooling on Growth, Income Distribution and Welfare
1.92	Chaim Fershtman	Privatization of a Firm in Oligpolistic Market (Hebrew)
2.92	Chaim Fershtman Arye Bebchuk	The Effects of Insider Trading on Insiders' Effort in Good and Bad Times
3.92	Jeffrey Church Neil Gandal	Integration, Complementary Products and Variety
4.92	Arthur Fishman	The Durability of Information, Market Efficiency and the Size of Firms: Search with Repeated Transactions

---

*The list of the Sackler Institute recent working papers follows this list. Copies of the working papers or a complete working paper list of the two Institutes can be obtained from: Mrs.Stella Padeh, The Foerder Institute for Economic Research, Tel–Aviv University, Tel–Aviv, 69978 Israel. Fax: 972–3–640–9908

5.92	Michele Piccione Ariel Rubinstein	Finite Automata Play a Repeated Extensive Game
6.92	Gil Bufman Leonardo Leiderman	Currency Substitution under NonExpected Utility Some Empirical Evidence
7.92	Arthur Fishman Neil Gandal	Experimentation in a Duopoly
8.92	Oz Shy	A Welfare Evaluation of Anti–Dumping Rules
9.92	Gideon Fishelson	Mortgage Market in Israel: 1950–1990 (Hebrew)
10.92	Arthur Fishman	The Informative Role of Prices in Markets with Endogenously Informed Buyers
11.92	Efraim Sadka	International Capital Flows: Efficiency of Savings and Investments
12.92	Elhanan Helpman	Innovation, Imitation, and Intellectual Property Rights
13.92	Zvi Eckstein and Kenneth I.Wolpin	Duration to First Job and the Return to Schooling: Estimates from a Search–Matching Model.
14.92	Adam B.Jaffe Manuel Trajtenberg Rebecca Henderson	Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations
15..92	Manuel Trajtenberg Rebecca Henderson Adam B. Jaffe	Ivory Tower Versus Corporate Lab: An Empirical Study of Basic Research and Appropriability
16.92	Timothy F.Bresnahan Manuel Trajtenberg	General Purpose Technologies: "Engines of Growth?"
17.92	Elhanan Helpman Leonardo Leiderman	Israel's Exchange Rate Band
18.92	Neil Gandal	Hedonic Price Indexes for Spreadsheets and an Empirical Test of the Network Externalities Hypothesis
19.92	Eitan Berglas David Pines	Transaction Costs in a Tiebout Setup
20.92	Gideon Fishelson	Endowment Effect and Market Effect: Revisited
21.92	Gene M.Grossman Elhanan Helpman	Protection for Sale
22.92	Eyal Sulganik	On the Structure of Blackwell's Equivalence Classes Of Information Systems

THE SACKLER INSTITUTE OF ECONOMIC STUDIES
List of Recent Working Papers

11.91	Oz Shy	International Standardization and Protection
12.91	Chaim Fershtman Arthur Fishman	Random Liquidity, Price Dispersion and Inflation
1.92	Edi Karni David Schmeidler	On the Uniqueness of Subjective Probabilities
2.92	Neil Gandal Nathan Sussman	Tickling the Tolerance: Incentives for Producing Low Quality Coins in French Medieval Mints
3.92	Yoram Weiss	Inflation and Price Adjustment: A Survey of Findings from Micro–Data
4.92	Ariel Rubinstein Asher Wolinsky	A Remark on Infinitely Repeated Extensive Games
5.92	Assaf Razin Shula Pessach	Targeting the Exchange Rate Under Inflation
6.92	Gil Bufman Leo Leiderman	Simulating an Optimizing Model of Currency Substitution
7.92	Ariel Rubinstein	On the Interpretation of Two Theoretical Models of Bargaining
8.92	Gideon Fishelson	Backstop Technology for an Exhaustible Resource: A New Look at an Old Problem
9.92	Gideon Fishelson	Imperfect Competition in a Market of an Exhaustible Resource
10.92	Oz Shy	A Welfare Evaluation of Comparison Advertising
11.92	Assaf Razin Efraim Sadka	International Migration and International Trade
12.92	Aba Schwartz	Seniority, Tenure and the Efficiency Wage Paradigm
13.92	Zvi Hercowitz	Macroeconomic Implications of Investment–Specific Technological Change
14.92	Elchanan Ben–Porath	Rationality, Nash Equilibrium and Backward Induction in Perfect Information Games