Learning by Forgetful Players

Hurkens, S.

Publication date:
1994

CentER for Economic Research

Discussion paper

[redacted]

Center
for
Economic Research

No. 9437

LEARNING BY FORGETFUL PLAYERS: FROM PRIMITIVE FORMATIONS TO PERSISTENT RETRACTS

by Sjaak Hurkens

May 1994

ISSN 0924-7815

Learning by forgetful players: From primitive formations to persistent retracts

Sjaak Hurkens*†

April 1994

Abstract

A product set of pure strategies is said to be closed under best replies if all best replies against all possible mixtures of these strategies are contained in the set. Minimal sets with this property are called minimal curb sets. It has been argued informally that the concept of minimal curb sets has an evolutionary flavour. In this paper we present a formal foundation to support this idea.

We construct a learning process that has two main characteristics: Players have a bounded memory and they play best replies against beliefs, formed on the basis of strategies used in the recent past. It is shown that this learning process leads the players to playing strategies from a minimal curb set. Moreover, this result continues to hold in the presence of mimickers and sophisticated players. When players are uncertain the process does not converge to a minimal curb set but to related solution concepts as curb*, robust or persistent sets, depending on how the uncertainty is modelled.

---

*CentER, Tilburg University P.O. Box 90153, 5000 LE Tilburg, The Netherlands.

†I would like to thank Peter Borm, Eric van Damme, Jürgen Eichberger, Drew Fudenberg and participants of the IIASA workshop on Evolutionary Game Dynamics in Biology and Economics for helpful discussions and comments. This research was sponsored by the Foundation for the Promotion of Research in Economic Sciences, which is part of the Netherlands Organization for Scientific Research (NWO).

1 Introduction

A product set of pure strategies is said to be closed under best replies if all best replies against all possible mixtures of these strategies are contained in the set. Minimal sets with this property are called minimal curb sets (Basu and Weibull (1991)). Curb sets are closely related to the better known persistent retracts. Kalai and Samet (1984) showed that every game has at least one persistent retract and that every persistent retract contains at least one (proper) Nash equilibrium. This enabled them to introduce the persistent equilibrium as a refinement of the Nash equilibrium concept.

Both concepts have been used in the literature. Kalai and Samet (1985) used persistency to achieve efficiency in unanimity games that are repeated as long as no agreement is reached. Blume (1993a) used the persistent retract as a set-valued solution concept in sender receiver games. Blume (1993b) shows that equilibria in minimal curb sets sometimes select the preferred outcome in one-sided cheap talk games. Hurkens (1993) shows that minimal curb sets always select the preferred outcome in games where several players have the possibility to send costly messages. Van Damme and Hurkens (1993) applied the concepts of curb and persistency in games of endogenous timing and Balkenborg (1993) did so in finitely repeated games.

In most of these papers it is argued informally that the concepts of curb and persistency have an evolutionary flavour. However, few or no attempts have been made to support this idea with an evolutionary foundation of the concepts.

We construct a learning process that has the following characteristic: Players have a bounded memory. On the basis of strategies played in the recent past, they form expectations about the strategies the other players will use, and best respond to these expectations. We assume that every period players are drawn from a heterogeneous pool. Different players may have different beliefs and therefore they may choose different actions. It is shown that, if the memory is long enough, play will settle down in minimal curb sets.

In some respects our results are stronger than those obtained thus far in the literature on learning. First, the process always converges.[1] Second, the set of curb stategies is a subset of the set of rationalizable strategies (Bernheim (1984) and Pearce (1984)). Hence, our learning process reduces the number of "plausible" strategies. This is in contrast with Milgrom and Roberts (1991) who show that a sequence that is consistent with adaptive learning will eventually lie within the set of serially undominated strategies, which is a superset of the set of rationalizable strategies. Third, it is reasonably simple to calculate the minimal curb sets of a game. It is not necessary to simulate the learning process in order to determine them.

From the main and basic theorem we derive several results for learning processes where players learn in a somewhat different way. Play still settles down in minimal curb sets when some players do not play best responses to past play, but are more sophisticated than that, or, on the contrary, are less sophisticated. If we allow players to have beliefs as if the other players in the game correlate their actions, play settles down in a primitive formation (Harsanyi and Selten (1988)), a variant of a minimal curb set. When players are uncertain, the process does not converge to a curb set but to related solution concepts as curb*, robust or persistent sets, depending on how the uncertainty is modelled. The learning processes presented in this paper may give the reader some insight in the differences and similarities between these related concepts. We also characterize two classes of games where our results go through, even if the players only observe the outcomes of past play, instead of the full descriptive strategies.

The rest of the paper is organized as follows. In section 2 we introduce some preliminaries concerning Markov chains and curb sets. Section 3 describes the model of learning as a Markov chain. Section 4 contains the main result: the ergodic sets of the Markov chain correspond one-to-one to the minimal curb sets of the underlying game. In sections 5 and 6 the above mentioned variations of the learning process are considered. In section 7 we consider the possibility that players make mistakes with small probability. Section 8 concludes.

---

[1]The process converges to a set. Within the set the process may continue to "drift".

2 Preliminaries

Let $G = (S_1, \ldots, S_n, u_1, \ldots, u_n)$ be a finite game with player set $N = \{1, \ldots, n\}$. Let $S = \prod_{i=1}^n S_i$ and $S_{-i} = \prod_{j \neq i} S_j$. For any finite set X let $\Delta(X)$ denote the set of probability distributions over X. For a distribution $\mu \in \Delta(S)$ let $\mu_i \in \Delta(S_i)$ be the marginal on S_i, and let $\mu_{-i} \in \Delta(S_{-i})$ be the marginal on S_{-i}, i.e.

$$\begin{aligned} \mu_i(s_i) &= \textstyle\sum_{s_{-i} \in S_{-i}} \mu(s_i, s_{-i}) && (s_i \in S_i) \\ \mu_{-i}(s_{-i}) &= \textstyle\sum_{s_i \in S_i} \mu(s_i, s_{-i}) && (s_{-i} \in S_{-i}) \end{aligned}$$

Of special interest are the probability distributions whose marginals on $S_1, \ldots, S_n$ are independent. The sets of these probability distributions will be denoted by Σ and Σ_{-i}, respectively. Although they are formally not the same, we will identify Σ with $\prod_{i=1}^n \Delta(S_i)$ and Σ_{-i} with $\prod_{j \neq i} \Delta(S_i)$ and trust that no confusion will result.

For $\mu \in \Delta(S)$ and $i \in N$ we let $\mathrm{BR}_i(\mu_{-i})$ denote the set of pure best replies against μ_{-i}. Let $\mathrm{BR}(\mu) = \prod_{i=1}^n \mathrm{BR}_i(\mu_{-i})$. For $F \subset \Delta(S)$ let $\mathrm{BR}_i(F) = \cup_{\mu \in F} \mathrm{BR}_i(\mu_{-i})$ and $\mathrm{BR}(F) = \prod_{i=1}^n \mathrm{BR}_i(F)$.

Definition 1 . *A non-empty cartesian product set $C = \prod_{i=1}^n C_i \subset S$ is said to be closed under best replies (or C is a curb set) if $BR(\prod_{i=1}^n \Delta(C_i)) \subset C$. Such a set is called a minimal curb set if it does not properly contain a curb set. Strategies contained in minimal curb sets are called curb strategies.*

It is straightforward to show that $\mathrm{BR}(\prod_{i=1}^n \Delta(C_i)) = C$ for any minimal curb set C. The notion of curb sets was introduced by Basu and Weibull (1991). Curb is mnemonic for closed under rational behaviour.

A strict Nash equilibrium is a curb set as a singleton. Strict Nash equilibria have almost all desired properties one can hope for, except existence. A lot of these properties carry over to minimal curb sets. For instance, every curb set contains the support of a proper equilibrium. Moreover, every game has at least one minimal curb set.

Minimal curb sets can be viewed as a set-valued generalization of strict equilibria: When an outsider recommends to all players to play strategies from a minimal curb set C, then all players will follow this recommendation if they expect the other players to do so. The comparison with strict equilibria is not completely valid: minimal curb sets may contain weakly dominated strategies.

Before we go further let us consider some examples where minimal curb sets have some cutting power.

Example A. Let G be given by the following normal form.

	L	R
T	2,2	0,0
B	0,0	1,1

Figure 1.

This is a pure coordination game. Since (T, L) and (B, R) are strict equilibria it is easy to see that $\{(T, L)\}$ and $\{(B, R)\}$ are minimal curb sets, and that there are no other ones. In particular, the support of the mixed equilibrium is not contained in any minimal curb set.

Example B. Let G be given by

	L	R
O	2,2	2,2
T	3,1	0,0
B	0,0	1,3

Figure 2.

G is the normal form representation of the extensive form game where player 1 has the choice between an outside option O which gives both players a payoff of 2, and entering a "battle of the sexes" game with player 2. This game has a unique minimal curb set, namely $R = \{(T, L)\}$.

These two examples are nice because the minimal curb sets are singletons, and hence consist of one strict Nash equilibrium. The following example is different.

Example C. Let G be given by the normal form in Figure 3.

	L	R
T	9,5	1,4
B	4,4	7,7

Figure 3.

Hurkens (1993) analyzes situations where some players can send a costly message to all players before a game is played. Suppose that player 1 can send one of two messages, m^0 or m^1, to player 2 before G is played. Suppose that it costs player 1 i units to send m^i. Let ma denote player 1's strategy "I send message m and choose action a" and let a^0a^1 denote player 2's strategy "I choose action a^i if I receive message m^i". Then the (reduced) normal form of the game with pre-play communication is as follows.

	LL	LR	RL	RR
m^0T	9,5	9,5	1,4	1,4
m^0B	4,4	4,4	7,7	7,7
m^1T	8,5	0,4	8,5	0,4
m^1B	3,4	6,7	3,4	6,7

Figure 4.

Now it can be checked that $\{m^0T\} \times \{LL, LR\}$ is the unique minimal curb set of this extended game. The set is not a singleton but it consists only of equilibria that involve sending the cheapest message and then playing the equilibrium preferred by player 1. In Hurkens (1993) similar results are obtained for a whole class of games with n players among which k have the possibility to send a costly message.

In the next section we will describe the learning process by means of a Markov chain. Therefore we will need some basic notions from the theory of Markov chains.

A finite stationary Markov chain is characterized by a pair (X, P), where X is a finite state space and $P : X \times X \to [0,1]$ is a transition matrix. The interpretation is that $P(x, x')$ is the probability that the process will move from x to x' in one period. We will denote $x \rightsquigarrow x'$ if there exist $k \in \mathbb{N}, x_0, \ldots, x_k \in X$ with $x_0 = x$, $x_k = x'$ and $P(x_i, x_{i+1}) > 0$ $(i = 0, \ldots, k-1)$. Now $\rightsquigarrow$ defines a weak order on X. Hence, we can define an equivalence relation on X:

$$x \sim y \quad \Leftrightarrow \quad x \rightsquigarrow y \text{ and } y \rightsquigarrow x$$

Let $[x]$ denote the equivalence class that contains x and let $Q = \{[x] | x \in X\}$ denote the set of equivalence classes. We define a partial order $\preceq$ on Q.

$$[x] \preceq [y] \quad \Leftrightarrow \quad y \rightsquigarrow x$$

The minimal elements with respect to the order $\preceq$ are called ergodic sets. The other elements are called transient sets. If the process leaves a transient set it can never return to that set. And if the process is in an ergodic set it can never leave this set. The elements of these sets are called ergodic and transient states. We have the following theorem.

Theorem 1 . *In any finite Markov chain, no matter where the process starts, the probability after k steps that the process is in an ergodic state tends to 1 as k tends to infinity.*

Proof. See e.g. Kemeny and Snell (1976).

3 The learning process

According to the Bayesian approach, a player forms some expectation about the strategies that will be played by the other players, and best responds to his expectation. How these expectations are formed is not clear. When the same game has been played before, possibly by different people, it seems reasonable to suggest that expectations are formed on the basis of information on past play. One way of using this information is to assume that a player's belief corresponds to the empirical frequency of strategies used in the past. This way of forming beliefs, known as fictitious play (Brown (1951) and Robinson (1951)), makes perhaps sense in matching models, but it is certainly not the only possible way of forming beliefs. One drawback of fictitious play is that it assumes that all people always form expectations in the same way. This implies that if different people have the same information, they will form the same beliefs and consequently they choose the same action. One can create some stochastic variability in the process by assuming that people only draw an incomplete sample of the information, as in Young (1993). There it is assumed that players learn how the game was played in m out of the most recent K periods. The players use a fictitious play rule to map samples into beliefs, and best respond to these beliefs. The great technical advantage of Young's approach is that the learning process can be described by a finite Markov chain on the state space $H = S^K$, consisting of all sequences of length K drawn from S. In order to determine the ergodic sets of such Markov chains, one needs only to specify which transitions occur with positive probability, and which occur with zero probability.

We will also describe a learning process by means of a finite Markov chain, but we need more variability in the responses of the players. In fact, we need the degree of variability that is present in Milgrom and Roberts' (1991) definition of adaptive play.[2]

Let $G = (S, u)$ be an n-person normal form game. Fix a positive integer K. Suppose we have a finite population of individuals that is partitioned into non-empty classes $V_1, \ldots, V_n$. The members of V_i are candidates to play role i in the game, and they all have the same payoff function u_i. Let $t = 0, 1, 2, \ldots$ denote succesive time periods. Game G is played once every period. In period t one individual is drawn from each class V_i. These individuals are going to play the appropriate roles in the game this period. We will refer to the individual that is drawn from V_i to play the game in the current period as player i, although the identity of this player may change from time to time.

Let $s(0), s(1), \ldots, s(t-1)$ denote the strategy profiles played up to period t. Player i receives some, but not necessarily all, information about play in the recent K periods, denoted by $I[s(t-K), \ldots, s(t-1)]$. Then he chooses a pure strategy $s_i(t)$ according to some rule. We will define below what kind of information a player may receive, and how he chooses a strategy as a function of this information. Then the players are put back in their class (or they die and are replaced by a new individual with the same utility function). This ends period t and we move up to period $t+1$. Again, from each class one individual is drawn to play the game. Player i receives some information $I[s(t+1-K), \ldots, s(t)]$ and chooses a strategy according to the same rule.

Since we will assume that all the rules are time-independent, this learning process can be described by a stationary Markov chain on the state space $H = S^K$. Call $\hat{h} \in H$ a successor of $h \in H$ if $\hat{h}$ is obtained from h by deleting the left most element and by adding some element $s \in S$ to the right. Let $r(\hat{h})$ denote the right most element

[2]See Section 8 for a comparison between the present paper and Milgrom and Roberts (1991).

of $\hat{h} \in H$. For $h = (s^{-K}, \ldots, s^{-1}) \in H$ let $\pi_i(h) = \{s_i^{-K}, \ldots, s_i^{-1}\}$ denote the set of strategies played by player i in the recent past. We will assume that our learning process is described by a transition matrix $P \in \mathcal{P}$, where $\mathcal{P}$ is defined as follows.[3]

Definition 2 .

Let $\mathcal{P}$ denote the set of transition matrices P, that satisfy for all histories $h, \hat{h} \in H$,

$$P(h, \hat{h}) > 0 \quad \Leftrightarrow \quad \begin{cases} \hat{h} \text{ is a successor of } h \\ r_i(\hat{h}) \in BR_i(\mu^i) \text{ for some } \mu^i \in \prod_{j \neq i} \Delta(\pi_j(h)) \quad (\text{all } i) \end{cases}$$

We will give two interpretations of a learning process that is described by some $P \in \mathcal{P}$. The first interpretation is close to the model of Young (1993). Fix a positive integer L. Before player i chooses a strategy in period t, he receives information about how the game was played by player j in the recent past, for all $j \neq i$. He receives L draws with replacement from the set $\{s_j(t-K), \ldots, s_j(t-1)\}$. A way of thinking about this sampling procedure is that player i passively hears about L precedents concerning the way player j played the game before. But player i is unaware of the fact that he might hear about the same precedent several times. Assume that all draws are independent, but more importantly, assume that each combination of draws occurs with positive probability. Player i's belief about the behaviour of player j corresponds to the empirical frequency of strategies in the sample of size L. Hence, this belief is one of a finite number of possible probability distributions. Namely, let $h = (s(t-K), \ldots, s(t-1))$ denote the recent history and let $\pi_j(h) = \{s_j(t-K), \ldots, s_j(t-1)\}$ denote the set of strategies played by player j in the recent past. Now player i's belief about player j's behaviour is contained in the set

$$B_j(h, L) = \{\mu_j \in \Delta(\pi_j(h)) | \mu_j(s_j) = l/L \text{ for some } l \in \{0, 1, \ldots, L\}\}.$$

We call the set $B^i(h, L) = \prod_{j \neq i} B_j(h, L)$ the L-grid distribution space for i induced by h. This learning process could be described by a transition matrix P' with the following properties.

[3] A transition matrix describes a learning process for a fixed game, G, and a fixed length of the memory, K. We will however suppress superscripts G and K.

(1) If $\hat{h}$ is not a successor of h, then $P'(h,\hat{h}) = 0$.

(2) If $\hat{h}$ is a successor of h, and s is the right most element of $\hat{h}$, then $P'(h,\hat{h}) > 0$ if and only if, for all i, $s_i \in \mathrm{BR}_i(\mu^i)$ for some $\mu^i \in B^i(h, L)$.

Note that as L increases, the grid becomes finer and finer, and the stochastic variability of the process increases. It seems that as L increases, P' "approaches" some $P \in \mathcal{P}$. There exists a 'generic' class of games for which it suffices, for the purpose of this paper, to choose L sufficiently large. However, in general we need a little bit more and therefore we assume that our learning process is described by some $P \in \mathcal{P}$.

Another interpretation of a learning process that is described by a transition matrix $P \in \mathcal{P}$ is the following. Suppose that the individuals in a class have different personal characteristics: They use the information on past play to know which strategies will certainly not be used (namely the ones that have not been played in the recent history). But each individual makes his own personal assessment of the probabilities with which the remaining strategies will be played. Some people are very optimistic and expect the best, while others are very pessimistic and expect the worst. And there will be a lot who have some intermediate beliefs. Of course, we need sufficient diversity in the different classes when this learning process is to be described by some $P \in \mathcal{P}$. Note, however, that this does not necessarily mean that these classes are large. Suppose that for each strategy $s_i \in S_i$, there is some individual in V_i who plays s_i, whenever it is a best reply to some belief that puts positive weight only on strategies that were played recently. (And he chooses a best reply to the most recent strategy otherwise.) Then we only need $|S_i|$ individuals in class V_i.

In the next section we will state and prove the main theorem of this paper: Play will settle down in minimal curb sets.

4 Ergodic sets

Fix $K \in \mathbf{N}$ as the length of the histories. Recall from section 2 that $h \rightsquigarrow \hat{h}$ means that there exist $k \in \mathbf{N}$, $h^0, \dots, h^k \in H = S^K$ such that $h^0 = h$, $h^k = \hat{h}$ and $P(h^i, h^{i+1}) > 0$. Now $\rightsquigarrow$ defines a weak order on H and hence we can define an equivalence relation on H and an order on the set of equivalence classes of H. We will be interested in the minimal elements of this order, the ergodic sets.

Let C be a minimal curb set of $G = (S, u)$. We say that $h \in H$ is a C-history if $h \in C^K$. We call h a curb history if it is a C-history for some minimal curb set C.

Now we are ready to state the main theorem.

Theorem 2 . *There exists $\underline{K} \in \mathbf{N}$ such that for all $K \geq \underline{K}$ and every Markov chain with a transition matrix $P \in \mathcal{P}$*

(i) If $Z \subset H$ is an ergodic set then $Z \subset C^K$ for some minimal curb set C.

(ii) For every minimal curb set C there exists exactly one subset $Z \subset C^K$ that is ergodic.

The theorem states that, if the history is long enough, any ergodic set is a set of C-histories for some minimal curb set C and that the set of C-histories contains one ergodic set. Hence, the ergodic states are curb histories. From Theorem 1 then the following corollary follows.

Corollary 1 . *The probability that the players are playing a curb strategy profile after k steps of the learning process tends to 1 as k tends to infinity, if histories are sufficiently long.*

The intuition for the theorem is quite clear. By having a large enough memory, players may have beliefs with large supports. This means that best replies against all kinds of mixtures will be played now and then. This creates so much stochastic variability that players sooner or later will play curb strategies. When they keep drawing the "right" samples, they will keep best responding against curb strategies, and hence they will play curb strategies again. It might happen that they will do this K periods in a row. The probability that this happens at a specific point in time is only small, but with probability one it will happen eventually. By that time all non-curb strategies will be forgotten. The strategies that will be played from that point on, will depend on the sample drawn, but it is sure that it will be curb strategies again.

Before we start with the actual proof of Theorem 2 we introduce some notation and state a lemma.

Let F be a non-empty subset of S. We define the projection of F on S_i as $p_i(F) = \{f_i | f \in F\}$ and we define $\text{span}(F) = \prod_{i=1}^{n} p_i(F)$. Hence, $\text{span}(F)$ is the smallest cartesian product set in S that contains F. Similarly, for a history $h = (s^{-K}, \ldots, s^{-1})$ we define $\pi_i(h) = \{s_i^{-K}, \ldots, s_i^{-1}\}$ and $\text{span}(h) = \prod_{i=1}^{n} \pi_i(h)$. We say that $B \subset S$ spans F if $\text{span}(B) = \text{span}(F)$.

For a history h let $\mathcal{B}^{ind}(h) = \{\mu \in \Sigma | \text{supp}(\mu) \subset \text{span}(h)\}$. This set contains all independent beliefs a Bayesian player might have when the process is in state h. Similarly, we define for a set $F \subset S$, $\mathcal{B}^{ind}(F) = \{\mu \in \Sigma | \text{supp}(\mu) \subset \text{span}(F)\}$. Let $M = \max_i |S_i|$.

Lemma 1 . *Let $s^1, \ldots, s^T \in S$ be such that $s^{t+1} \notin \text{span}(\{s^1, \ldots, s^t\})$ for all $t = 1, \ldots, T-1$. Then $T \leq \sum_{i=1}^{n} |S_i| - (n-1)$.*

Proof. Easy and hence omitted. □

Proof of Theorem 2. Take $\underline{K} = \sum_{i=1}^{n} |S_i| - (n-1) + M$ and let $K \geq \underline{K}$. Let $P \in \mathcal{P}$.

Let $h^t = (x^{K-t}, \ldots, x^1, s^1, \ldots, s^t)$ be a particular history and assume that $F^t = \text{span}(\{s^1, \ldots, s^t\})$ is not a curb set. Then there exists $s^{t+1} \in \text{BR}(\mathcal{B}^{ind}(F^t)) \setminus F^t$. Let $h^{t+1} = (x^{K-t-1}, \ldots, x^1, s^1, \ldots, s^{t+1})$. Then $P(h^t, h^{t+1}) > 0$. Starting from an arbitrary history h^1 we can apply this argument repeatedly. By Lemma 1 we know that there exists $T \leq \underline{K} - M$ such that $h^1 \rightsquigarrow h^T = (x^{K-T}, \ldots, x^1, s^1, \ldots, s^T)$ and such that $F^T = \text{span}(\{s^1, \ldots, s^T\})$ is a curb set. Let $C \subset F^T$ be a minimal curb set and let $\{b^1, \ldots, b^M\}$ span C. Since every strategy in a minimal curb set is a best reply to some belief concentrated on this set and since $K \geq M + T$, we have $h^T \rightsquigarrow (\ldots, s^1, \ldots, s^T, b^1, \ldots, b^M) \rightsquigarrow (b^1, \ldots, b^M, b^M, \ldots, b^M)$.

The above shows that for every history h, there exists a minimal curb set C such that for every set $\{b^1, \ldots, b^M\}$ that spans C, we have $h \rightsquigarrow (b^1, \ldots, b^M, b^M, \ldots, b^M)$. Furthermore, the definition of $\mathcal{P}$ implies that if h is a C-history and $h \rightsquigarrow \hat{h}$, then $\hat{h}$ is also a C-history.

The second observation implies that the set of C-histories contains an ergodic set, for any minimal curb set C. The first observation then implies that the set of C-histories contains exactly one ergodic set, and that there are no other ergodic sets. □

It is not true in general that every curb history is an ergodic state. This is so because not every curb history can be reached from any other curb history.[4] Consider the game in Figure 5.

	a_2	b_2	c_2
a_1	4,1	1,4	2,3
b_1	1,4	4,1	2,3
c_1	3,2	3,2	0,0

Figure 5.

This game has only one curb set, namely the set of all pure strategy combinations. But the profile $\bar{h} = (c, c, \ldots, c)$ cannot be reached under the learning process from any other history. This is so because c is only a best reply against some mixtures of a and b. Hence, there exists no h with $P(h, \bar{h}) > 0$.

We certainly do not claim that the lower bound on K that was given in the proof of Theorem 2 is sharp. The example in Figure 6 shows however that memories must not be too short.

	a	l	c	r
A	4,4	2,2	2,2	2,2
T	2,2	5,0	0,5	0,0
M	2,2	0,0	5,0	0,5
B	2,2	0,5	0,0	5,0

Figure 6.

It is not difficult to see that if $K = 2$, then the set of histories $\{(s^{-2}, s^{-1}) | s^{-j} \in \text{span}(\{Tl, Mc, Br\})\}$ contains an ergodic set. Take for example the history (Tl, Mr).

[4]However, from the proof it follows easily that if $s \in C$ for some minimal curb set C, then there exists an ergodic state h with $r(h) = s$. Hence, every strategy in C will be played infinitely often once the ergodic set contained in C^K is entered.

Agents from pool V_1 will play a best reply against $\alpha l + (1-\alpha)r$, for some $\alpha \in [0,1]$. Hence, they will play T or B. But the unique curb retract is the singleton $\{(A,a)\}$. So the history must not be too short. Note that if $K = 3$ and the process is in state (Tl, Mr, Mc), then there will be some agent in V_1 who will play A, since A is the best reply against $\frac{1}{3}l + \frac{1}{3}c + \frac{1}{3}r$.

Note that the game from Figure 6 has a unique equilibrium, namely (A,a). This equilibrium is strict. Since every curb set contains the support of a Nash equilibrium and since a strict equilibrium forms a curb set as a singleton, it follows that this game has a unique minimal curb set. Hence, if players behave as described by our learning process then they will eventually play the equilibrium. This reasoning holds for all games that have a unique equilibrium that happens to be strict. So we proved

Corollary 2 . *Suppose that s is the unique Nash equilibrium of G and $BR(s) = \{s\}$. The probability that players are playing the equilibrium after k steps of the learning process tends to 1 as k tends to infinity, if histories are sufficiently long.*

5 Variations on the same theme

We remarked before that one only needs to know which entries of the transition matrix are positive and which are zero in order to characterize the ergodic sets. In the proof of Theorem 2 we used that certain entries are positive (together with Lemma 1) to show that the process can move from any history h to a curb history $\hat{h}$ in a finite number of periods. Furthermore, we used the fact that certain entries are zero to ensure that the process can not leave the set of C-histories, for any curb set C.

It is possible to prove Theorem 2 for an even bigger class of transition matrices. Let $P \in \mathcal{P}$ and let $\tilde{P}$ be a transition matrix that satisfies, for any minimal curb set C,

$$P(h,\hat{h}) > 0 \;\Rightarrow\; \tilde{P}(h,\hat{h}) > 0 \tag{5.1}$$

$$h \in C^K \text{ and } \tilde{P}(h,\hat{h}) > 0 \;\Rightarrow\; \hat{h} \in C^K \tag{5.2}$$

Let $\tilde{\mathcal{P}}$ denote the set of all such transition matrices. It is obvious that Theorem 2 holds for all $P \in \tilde{\mathcal{P}}$. We will consider two subsets of $\tilde{\mathcal{P}}$, namely $\mathcal{P}^{soph}$ and $\mathcal{P}^{mim}$. The transition matrices in these sets correspond to learning processes where some players are more sophisticated (in the case of $\mathcal{P}^{soph}$) or less sophisticated (in the case of $\mathcal{P}^{mim}$). It turns out that for these two classes we can prove slightly stronger results.

5.1 More and less sophisticated players

Suppose that not all individuals in the classes are Bayesian players, but that some individuals are mimickers. Mimickers don't form expectations but just observe how other agents in the same role have played the game during (some of) the last K periods. Then they choose one of these strategies at random. When we retain our assumption about the Bayesian players, this learning process can be described by a transition matrix $P \in \mathcal{P}^{mim}$, where $\mathcal{P}^{mim}$ is defined as follows.

Definition 3 .

Let $\mathcal{P}^{mim}$ denote the set of transition matrices P, that satisfy for all histories $h, \hat{h} \in H$,

$$P(h,\hat{h}) > 0 \quad \Leftrightarrow \quad \begin{cases} \hat{h} \text{ is a successor of } h \\ r_i(\hat{h}) \in BR_i(\mathcal{B}^{ind}(h)) \text{ or } r_i(\hat{h}) \in \pi_i(h) \quad (\text{all } i) \end{cases}$$

Obviously, $\mathcal{P}^{mim} \subset \tilde{\mathcal{P}}$, hence Theorem 2 holds for all $P \in \mathcal{P}^{mim}$. We can prove a slightly stronger result: Alle curb histories are ergodic states.

Theorem 3 . *There exists $\underline{K} \in \mathbb{N}$ such that for all $K \geq \underline{K}$ and for every Markov chain with a transition matrix $P \in \mathcal{P}^{mim}$, $Z \subset H$ is an ergodic set if and only if $Z = C^K$ for some minimal curb set C.*

Proof. Using the proof of Theorem 2, it suffices to show that if C is a minimal curb set and h and $\hat{h}$ are C-histories, then $h \rightsquigarrow \hat{h}$.

Let $\hat{h} = (s^{-K}, \ldots, s^{-1})$. We can choose a set $B = \{b^1, \ldots, b^M\}$ that spans C such that $s^{-j} \in \text{span}(\{b^1, \ldots, b^j\}$, for $j = 1, \ldots, M$. From the proof of Theorem 2 we know that $h \rightsquigarrow (b^1, \ldots, b^M, s^{-K}, \ldots, s^{-(M+1)}) =: \bar{h}$. Because of the special way we chose B (and because players sometimes mimic) we have $\bar{h} \rightsquigarrow \hat{h}$. □

It is possible to prove Theorem 3 with a smaller lowerbound on the length of the memory by making full use of the presence of the mimickers. We will not pursue that here. We just remark that for weakly acyclic games, the class of games considered in

Young (1993), we could take $\underline{K} = 1$.

The learning process we considered implies that Bayesian players play best responses against past play. If a player knew that other players are following this process, he could do better by playing a strategy that is a best reply against a strategy profile, consisting of best responses for the other players against past play. Of course, we may have players who foresee that others are going to play best responses to best replies to past play. We could have even more sophisticated players. When we assume that in a class many different levels of sophistication are represented, we have a learning process with sophisticated players. (See also Milgrom and Roberts (1991).)

Formally, let h be a particular history and let $T^0(h) = \text{span}(h)$. Define recursively $T^{j+1}(h) = \text{span}(T^j(h) \cup \text{BR}(\mathcal{B}^{ind}(T^j(h))))$. Since $T^{j+1}(h) \supset T^j(h)$ and S is finite, $T^\infty(h) = \text{span}(\cup_{j=0}^{\infty} T^j(h))$ is well-defined. Again, we define a whole set of transition matrices that correspond to learning processes with sophisticated players. We will denote this class by $\mathcal{P}^{soph}$, where $\mathcal{P}^{soph}$ is defined as follows.

Definition 4 .

Let $\mathcal{P}^{soph}$ denote the set of transition matrices P, that satisfy for all histories $h, \hat{h} \in H$,

$$P(h, \hat{h}) > 0 \quad \Leftrightarrow \quad \begin{cases} \hat{h} \text{ is a successor of } h \\ r(\hat{h}) \in BR(T^\infty(h)) \end{cases}$$

It is obvious that $\mathcal{P}^{soph} \subset \tilde{\mathcal{P}}$ and hence Theorem 2 is valid, also for this class. We can prove a stronger result: In the presence of sophisticated players we only need a memory of length one. The intuition for this result is that sophisticated players can do all the learning in their heads. They might foresee all the steps that needed to be executed in the case of no sophisticated players.

Theorem 4 . *For all $K \geq 1$ and all Markov chains with a transition matrix $P \in \mathcal{P}^{soph}$ we have $Z \subset H$ is an ergodic set if and only if $Z = C^K$ for some minimal curb set C.*

Proof. For notational convenience we just give the proof for $K = 1$. Now $H = S$ and we can define $T^\infty(s)$ for all $s \in S$. Note that $T^\infty(s)$ is a curb set and hence there exists a minimal curb set $\bar{C} \subset T^\infty(s)$. If $\bar{s} \in \bar{C}$ then $P(s, \bar{s}) > 0$.

Note that if $s \in C$ for some minimal curb set C then $T^{\infty}(s) = C$. Hence, if $s', s'' \in C$, then $P(s', s'') > 0$. □

The reader may have noticed that this sophisticated learning process has some similarities with the notion of rationalizability (Bernheim (1984) and Pearce (1984)). The difference is that rationalizability corresponds with a process of iterative elimination of strategies that are never best replies (starting with the whole space of strategy profiles) whereas our learning process implies the addition of best replies (starting from a history). The bounded memory of the players causes play to settle down in a *minimal* curb set.

The similarity of rationalizable and curb strategies has already been pointed out by Basu and Weibull (1991) and Balkenborg (1992): Call a set $C = \prod_{i=1}^{n} C_i$ tight if $\mathrm{BR}(\prod_{i=1}^{n} \Delta(C_i)) = C$. The maximal tight set is the set of rationalizable strategies, the minimal tight sets are just the minimal curb sets. In particular, every curb strategy is rationalizable.

5.2 Uncertain players

Consider the game from Figure 7.

	L	R
T	1,1	1,1
B	1,1	0,0

Figure 7.

This game has a unique curb set: it consists of all pure strategy profiles. When players behave as described by any of the learning processes they will regularly be playing (B, R)! This might seem a bit strange. It could not happen if the players were careful and only played undominated best replies. Then they would finally be playing only (T, L).

This example shows a drawback of the notion of minimal curb sets: They can contain strategies that are weakly dominated. Therefore let us introduce the notion of sets that are closed under undominated best replies. Formally, s_i is weakly dominated by s_i' if $u_i(s_i, s_{-i}) \leq u_i(s_i', s_{-i})$ for all s_{-i} with strict inequality for at least one s_{-i}. Let $\mathrm{UBR}(\sigma)$ denote the set of pure best replies against σ that are not weakly dominated.

Definition 5 . *A non-empty cartesian set $C = \prod_{i=1}^{n} C_i$ is closed under undominated best replies (or C is a curb* set) if for all $\sigma \in \prod_{i=1}^{n} \Delta(C_i)$, $UBR(\sigma) \subset C$. Such a set is called a minimal curb* set if it does not properly contain a set that is closed under undominated best replies. Strategies contained in minimal curb* sets are called curb* strategies.*

Lemma 2 . *Every curb set contains a curb* set. Every minimal curb* set contains the support of a Nash equilibrium. Curb* strategies are not weakly dominated.*

Proof. Easy and hence omitted. □

It is easy to adjust the learning process so that players will end up playing curb* strategies. Just replace 'best replies' by 'undominated best replies' and analogies of Theorems 2, 3 and 4 can be proved easily. On the level of Bayesian players this means that, although they have certain beliefs, they are not completely sure that these beliefs are "correct".[5] Therefore they should be careful and only play undominated best replies.

The approach taken above is a bit unsatisfactory since the uncertainty is not modelled. We will do that now. Remember the sampling procedure described in section 3. Every time an individual is drawn from class V_i, he hears about L precedents concerning the way player j played this game before. This sample is transformed (by the fictitious play rule) into a belief μ^i from the L-grid distribution space $B^i(h, L)$, where h denotes the recent history of plays.

Now suppose that the final belief of this player is not necessarily μ^i, but some $\hat{\mu}^i$ "close" to μ^i, reflecting the uncertainty of this player. This uncertainty may stem from the fact that the player realizes that he only draws a sample, and that μ^i is only a point estimate of the distribution of strategies. The final belief $\hat{\mu}^i$ could be a draw from some "confidence interval" around μ^i. This draw might depend on personal characteristics, as well as on other external factors. We will just assume that $\hat{\mu}^i$ is drawn from the uniform distribution over $B_\epsilon(\mu^i) = \{\sigma^i \in \Sigma_{-i} | d^i_{max}(\mu^i, \sigma^i) \leq \epsilon\}$, where $\epsilon > 0$ is fixed[6] and where

[5] The uncertainty of the players could stem from the fact that players may realize that other players have different samples. Anyway, sometimes players "are right" to be uncertain since it is possible that a history h is followed by the play of s, where $s \notin \text{span}(h)$.

[6] We could take $\epsilon = 1/L$ to reflect the intuition that bigger samples should result in smaller confidence intervals.

$d^i_{max}(\mu^i, \sigma^i) = \max_{s_{-i} \in S_{-i}} |\mu^i(s_{-i}) - \sigma^i(s_{-i})|$. Note that, for large L, the union of these intervals over all L-grid distibutions induced by h, consists of all probability distributions close to $\prod_{j \neq i} \Delta(\pi_j(h))$.

What consequences does this have for our learning process? Or, in other words, what strategies will be played with positive probability after each possible history? Well, let $h \in H$ and let $s_i \in S_i$. Before we had that s_i was played with positive probability, whenever there was some $\mu^i \in \prod_{j \neq i} \Delta(\pi_j(h))$ such that $s_i \in \mathrm{BR}_i(\mu^i)$. Now we have that s_i is played with positive probability, only if the stability region of s_i,

$$\mathrm{St}_i(s_i) = \{\sigma_{-i} \in \Sigma_{-i} | s_i \in \mathrm{BR}_i(\sigma_{-i})\},$$

has positive probability under the uniform distribution over $B_\epsilon(\hat{\mu}^i)$, for some L-grid distribution $\hat{\mu}^i$ induced by h. For sufficiently large L, this is equivalent to

$$\mu_{-i} \in \mathrm{cl}(\mathrm{int}(\mathrm{St}_i(s_i))), \tag{5.3}$$

for some $\mu_{-i} \in \prod_{j \neq i} \Delta(\pi_j(h))$, where $\mathrm{cl}(\cdot)$ and $\mathrm{int}(\cdot)$ stand for closure and interior (in the topological space Σ_{-i}), respectively.

Note that if $\mu_{-i} \in \mathrm{int}(\mathrm{St}_i(s_i))$, then s_i is a best reply against each strategy in an open neighbourhood of μ_{-i}. Up to equivalence, s_i is then also the unique (and undominated) best reply against this neighbourhood, and s_i is called a robust best reply against μ_{-i}. If only (5.3) is satisfied, there is some non-empty open set close to μ_{-i} against which s_i is the unique best reply, and we call s_i a semi-robust best reply against μ_{-i}, which is denoted by $s_i \in \mathrm{SRBR}_i(\mu_{-i})$. As opposed to robust best replies, semi-robust best replies always exist, and there may exist several semi-robust best replies against some μ_{-i}, even if player i has no equivalent strategies. It is easy to see that semi-robust best replies are not weakly dominated. Similar to the case with the (undominated) best reply correspondence we define

Definition 6 (Balkenborg (1992))

A non-empty cartesian set $C = \prod_{i=1}^n C_i$ is closed under semi-robust best replies (or C is a robust set) if $SRBR(\prod_{i=1}^n \Delta(C_i)) \subset C$. Such a set is called a minimal robust set if it does not properly contain a set that is closed under semi-robust best replies.

It is easy to see that every curb* set contains a robust set, but not every minimal robust set is (contained in) a minimal curb* set. Moreover, every robust set contains the support of a Nash equilibrium.

The learning process where players are uncertain can be described by a Markov chain that is very similar to the ones we had before. Just replace 'best replies' by 'semi-robust best replies' and analogies of Theorems 2, 3 and 4 can be proved easily. Play will settle down in a minimal robust set.

For 'generic' normal form games the minimal curb, curb* and robust sets coincide with the persistent sets. Persistent sets consist of the extreme points of persistent retracts (Kalai and Samet (1984)). As a matter of fact, for games in which no player has equivalent strategies, the minimal robust sets coincide with the persistent sets (see Balkenborg (1992)). However, many normal form games are interesting because they are the normal form representation of an extensive form game, and these are not 'generic' in the class of normal form games. This is due to the fact that there may be strategies in the extensive form game that preclude some information sets (or subgames) from being reached. This implies that curb sets may differ from robust sets. To illustrate this difference consider the following example that is taken from Hurkens (1993).

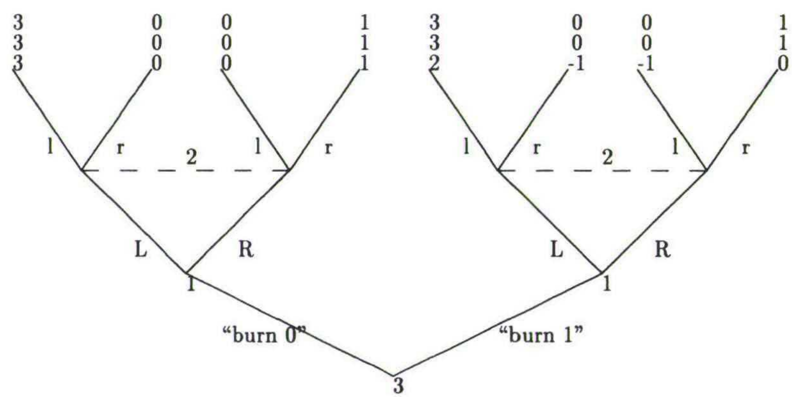

Figure 8.

Consider the game in Figure 8. Player 3 can decide to burn one unit before players 1 and 2 play a simultaneous move coordination game. Consider the strategy profile $s^{\text{ineff}} = (RR, rr, \text{"burn 0"})$. The singleton set containing this profile is a persistent and robust set: Consider the history $\bar{h} = (s^{\text{ineff}}, \ldots, s^{\text{ineff}})$. Player 3 has a unique best reply against s^{ineff}, namely "burn 0"; players 1 and 2 have a lot of (undominated) best replies against s^{ineff}, but in a small neighbourhood outside the set, that is when they are a little bit uncertain, they have a unique best reply. This is due to the fact that players 1 and 2 have an interest in choosing the same action: in a small neighbourhood player 1 plays 'R' with a very high probability, whether or not player 3 burnt something, and hence player 2 has to choose 'r', whether or not player 3 burnt something.

In contrast, the only minimal curb (or curb*) set of this game consists of all strategy profiles leading to the payoff vector $(3,3,3)$. When the system is in state $\bar{h}$, players 1 and 2 may change their choice of action in the subgame that is not reached, that is, when player 3 chooses to "burn 1" they may play L and l, respectively.

This example shows that the uncertainty we have introduced has a rather strange effect. By adding a little bit of uncertainty players are still quite certain about the strategies that will be used, but they are also certain that all information sets will be reached with positive probability. Therefore they have to play a best reply against the strategy profile that they believe to be played almost certainly, in all information sets, although many of these information sets will not be reached if this strategy profile is indeed played.

The peculiarity may very well be due to the fact that the game is an extensive form game, while curb is defined for normal form games. In our learning process we assumed that players know the *strategies* played in the past. For extensive form games it makes more sense to assume that players only observe the *outcomes*. We deal with this issue in section 6.

5.3 Dependent beliefs

Throughout this paper we assumed that a player's belief about the strategies of the other players is independent, i.e. is an element of Σ_{-i}. This was a consequence of the sampling procedure we described in section 3. Players receive information about the strategies of the players individually. Moreover, if players realize that the players are deciding simultaneously and independently, then it is natural to have independent beliefs. There are however two problems concerning the independency of beliefs.

First of all, do players indeed decide independently? After all, the choices of all players depend (via the samples) indirectly on the same recent history. History might act as a correlation mechanism. Secondly, our other interpretation of the learning process was that personal characteristics are important to form beliefs. All players expect that strategies that have not been played recently, will not be played, but different players may have different assessments of the probabilities with which the remaining strategies are played. In view of this interpretation, an individual from class V_i might have a dependent belief, i.e. an element of $\Delta(S_{-i})$. For instance, he might believe that the other players can correlate their strategies. It does not really matter whether or not the other players do correlate, what matters is that some individuals may believe that they do.

In this section we will examine the consequences of allowing players to have dependent beliefs. We will assume that the classes are very diverse: If h denotes the recent history and $s_i \in \mathrm{BR}_i(\mu^i)$ for some $\mu^i \in \Delta(\mathrm{span}(h))$, then s_i will be played with positive probability. Again, we will define a whole set of transition matrices describing such learning processes. Let $\mathcal{B}^{dep}(h) = \{\mu \in \Delta(S) | \mathrm{supp}(\mu) \subset \mathrm{span}(h)\}$ denote the set of all dependent beliefs a player may have.

Definition 7 .

Let $\mathcal{P}^{dep}$ denote the set of transition matrices P, that satisfy for all histories $h, \hat{h} \in H$,

$$P(h,\hat{h}) > 0 \quad \Leftrightarrow \quad \begin{cases} \hat{h} \text{ is a successor of } h \\ r(\hat{h}) \in BR(\mathcal{B}^{dep}(h)) \end{cases}$$

We can prove a theorem similar to Theorem 2. Of course, the process will in general not converge to a minimal curb set, but to a cartesian set $F = \prod_{i=1}^{n} F_i$ that is minimal with respect to the following property: If $\mu \in \Delta(F)$ and $s_i \in \mathrm{BR}_i(\mu_{-i})$, then $s_i \in F_i$. Following Harsanyi and Selten (1988) we call such a set a primitive formation.[7]

[7] Harsanyi and Selten (1988) consider this concept in the agent normal form.

Theorem 5 . *There exists $\underline{K} \in \mathbb{N}$ such that for all $K \geq \underline{K}$ and for every Markov chain with a transition matrix $P \in \mathcal{P}^{dep}$*

(i) If $Z \subset H$ is an ergodic set then $Z \subset F^K$ for some primitive formation F.

(ii) For every primitive formation F there exists exactly one ergodic subset $Z \subset F^K$.

We omit the proof because it is essentially the same as the proof of Theorem 2. We just have to observe that if F is a primitive formation and $s \in F$, then s_j is a best reply against some (dependent) belief concentrated on F.

Obviously, analogies of Theorems 3 and 4 to the case of dependent beliefs also exist. The same is true for the results of section 5.2 on undominated best replies and semi-robust best replies. Analogous to curb* and robust sets we could define primitive* and robust formations. The reader should be aware, though, that the definition of semi-robustness needs to be adapted. In the context of dependent beliefs we say that s_i is a semi-robust best reply against $\mu_{-i} \in \Delta(S_{-i})$ if $\mu_{-i} \in \mathrm{cl}(\mathrm{int}(\mathrm{St}_i(s_i)))$, where $\mathrm{cl}(\cdot)$ and $\mathrm{int}(\cdot)$ stand for closure and interior, respectively, in the topological space $\Delta(S_{-i})$, and where $\mathrm{St}_i(s_i) = \{\mu_{-i} \in \Delta(S_{-i}) | s_i \in \mathrm{BR}_i(\mu_{-i})\}$ is the stability region of s_i.

Of course, in a two person game the primitive formations are identical to the minimal curb sets. Moreover, every primitive formation contains a minimal curb set. Hence, if a game has a unique minimal curb set C which is also a primitive formation, then C is also the unique primitive formation. Similar statements can be made about the other concepts with the help of the following diagram. In this diagram $X \supset Y$ means that every X contains an Y, but not every Y is contained in an X.

$$\begin{array}{ccccc}
\text{primitive formation} & \supset & \text{primitive* formation} & \supset & \text{robust formation} \\
\cup & & \cup & & \cup \\
\text{min. curb set} & \supset & \text{min. curb* set} & \supset & \text{min. robust set}
\end{array}$$

Remark. Note that our definition of the transition matrices does not correspond to what one may call "correlated learning". Suppose that in a three player game player 3 observes that the other players played TL and BR in the last two periods. Then, under our assumption of dependent beliefs, it is possible that player 3 believes that TR and BL will be played, both with probability 1/2. One may feel that only beliefs of the form $\alpha TL + (1-\alpha)BR$ should be allowed. We do not know whether such "correlated learning" processes converge to some static set-valued solution concept.

6 Learning from outcomes

Throughout this paper we assumed that players know the strategies that were used in the past. This assumption is reasonable when the players in the underlying game choose their actions simultaneously. But if the underlying game is in fact an extensive form game, it makes more sense to assume that players only observe the outcomes, i.e. the paths in the tree generated by the strategies. Consider for example the "burning money" game in Figure 8. Suppose player 3 chose to "burn 0" in the last period. How could he know how players 1 and 2 would have reacted to "burn 1"? In fact, he can't, although he may have some beliefs.

In this section we will consider the case where players only observe the outcomes in the recent past. We assume that all agents form expectations on the basis of observed outcomes, and that different agents within a pool may form different beliefs. We pose only one restriction on the beliefs: When a player is able to conclude from the observed outcomes that a particular strategy has not been played during the last K periods, then he expects it will not be played next period. As before, we assume that the classes are very diverse: As soon as strategy s_i is a best reply against some independent belief, satisfying this restriction, then s_i will be played with positive probability.

We will define a class of transition matrices that correspond to such a "learning from outcomes" process, and we denote this class by $\mathcal{P}^{out}$. Before we can do so, we need some notation.

Let G be an extensive form game. Let $\mathcal{O}$ denote the set of outcomes (i.e. paths in the tree from the root to an endpoint) and let $o : S \rightarrow \mathcal{O}$ be the mapping that assigns to a pure strategy combination the outcome it generates. We will assume that there are no moves of Nature in G, since this mapping is not well-defined if there are. For a history $h = (s^{-K}, \ldots, s^{-1})$, let $\text{outc}(h) = \{o(s^{-K}), \ldots, o(s^{-1})\}$. Note that $\text{outc}(h)$ summarizes the information a player has. Let $\text{cons}_i(h) = \{s_i \in S_i | \exists s_{-i} \in S_{-i} \text{ s.t. } o((s_i, s_{-i})) \in$

outc(h)} denote the set of strategies of player i that are consistent with the observed outcomes. Let cons(h) $= \prod_{i=1}^{n} \text{cons}_i(h)$.

Definition 8 .

Let $\mathcal{P}^{out}$ denote the set of transition matrices P, that satisfy for all histories $h, \hat{h} \in H$,

$$P(h,\hat{h}) > 0 \quad \Leftrightarrow \quad \begin{cases} \hat{h} \text{ is a successor of } h \\ r(\hat{h}) \in BR(\mathcal{B}^{ind}(cons(h))) \end{cases}$$

In general, it is not true that play will settle down in minimal curb sets. Note that cons(h) $\supset$ span(h). This implies that if $P \in \mathcal{P}$ and $P(h,\hat{h}) > 0$, then $P^{out}(h,\hat{h}) > 0$ for all $P^{out} \in \mathcal{P}^{out}$. Using part of the proof of Theorem 2, it follows that, if K is large enough, for every history h and every $P^{out} \in \mathcal{P}^{out}$, there exists a curb history $\tilde{h}$ such that $h \rightsquigarrow \tilde{h}$. The problem is that there might exist a history $\hat{h}$, which is not a curb history, such that $\tilde{h} \rightsquigarrow \hat{h}$. This might even happen in 'generic' extensive form games, as the game from Figure 9 shows.

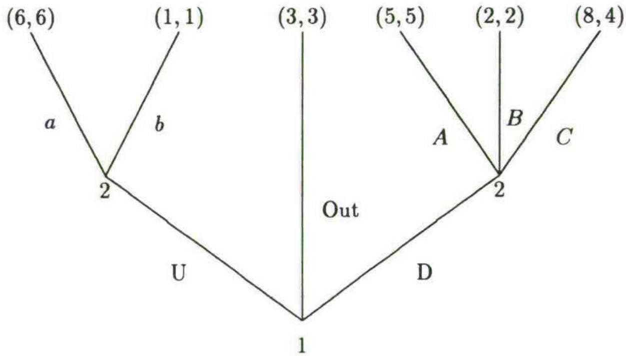

Figure 9.

This game has a unique minimal curb set, namely $\{U, D\} \times \{aA, aB, aC, bA\}$. However, suppose that in the recent (curb) history the strategy combinations (D, aB) and (U, bA) were played. Hence, player 1 observes (amongst other things) the outcomes DB and Ub. He might believe that the strategy bB was played, and will be played again next period. If he does so, he will choose 'Out', which is not a curb strategy.

The above example seems to suggest that there is no hope to obtain a result like Theorem 2 in the case of learning from outcomes. There are however two classes of games for which such an analogy does exist. The first class consists of the extensive form games without moves of nature, where each player has only one information set at which he has to make a choice. For obvious reasons we call such a game an agent normal form game without moves of nature, and we denote the class by ANF. The second class of games consists of those games G that have the property that any minimal curb set C of G corresponds to a single outcome, i.e. the set $\{o(c) | c \in C\}$ is a singleton. We denote this class by SCO (single curb outcome). Examples of these games are shown in Figures 2, 4 and 8.

To prove the above claims we just need to show that $\mathcal{P}^{out} \subset \tilde{\mathcal{P}}$, where $\tilde{\mathcal{P}}$ is as defined at the beginning of section 5. Part (5.1) follows from $\text{span}(h) \subset \text{cons}(h)$, part (5.2) follows from the next lemma.

Lemma 3 . *Let $G \in ANF$ or $G \in SCO$ and let C be a minimal curb set of G. Then*

$$h \in C^K \quad \Rightarrow \quad cons(h) \subset C$$

Proof. First consider the case $G \in$ ANF. Let j be a player. If there is an outcome $o(s^{-m}) \in \text{outc}(h)$ that does not intersect j's information set, then it follows that $\text{BR}_j(s^{-m}) = S_j$. This implies that $C_j = S_j \supset \text{cons}_j(h)$. If there is no such outcome, all outcomes intersect j's information set and $\text{cons}_j(h) = \pi_j(h) \subset C_j$. Hence, $\text{cons}(h) \subset C$.

Now consider the case $G \in SCO$. Let $\bar{s} = r(h)$. Now we have $\text{outc}(h) = \{o(\bar{s})\}$. Let j be a player and suppose $s_j \in \text{cons}_j(h)$. In any information set of j that intersects $o(\bar{s})$, s_j picks the same action as $\bar{s}_j$, since s_j is consistent with h. Since $G \in SCO$, we have that $\bar{s}_j$ is a best reply against $\bar{s}_{-j}$. But this implies that s_j is a best reply against $\bar{s}_{-j}$ as well, and hence $s_j \in C_j$. □

The reader can check that there are also analogies of Theorems 3 and 4 to the case where players learn from outcomes. The definiton of a mimicker needs to be adapted, since players don't observe strategies. We may assume that mimickers choose at random a strategy from the set of strategies that are consistent with (some of) the observed outcomes. There is also an analogy of Theorem 5, where players beliefs are not independent. There are however no analogues for the results of section 5.2 on the refined notions of undominated best replies or of semi-robust best replies. This is due to the fact that strategies that are consistent with a curb* history, may be weakly dominated. The game of Figure 9 shows an example of such a case: The only curb* strategy is (U, aA), but aB and aC are consistent with the curb* outcome.

7 Learning and experimentation

In many papers on learning, experimentation plays a prominent role. (See e.g. Kandori, Mailath and Rob (1993), Samuelson (1993), Young (1993) and Fudenberg and Kreps (1988)).

In Young (1993), Samuelson (1993) and Kandori et al. (1993) the possibility of experimentation (or mistakes, or mutations) implies that the Markov chain describing the learning process becomes irreducible, and hence has a unique stationary distribution. By taking the limit as the experimentation rate tends to zero, one stationary distribution of the unperturbed process is selected. In Young (1993) and Kandori et al. (1993) this yields typically a unique so called stochastically stable state because they consider a special class of games. Samuelson (1993) considers games with alternative best replies and then the support of the limit distribution consists usually of one or more line segments.

It turns out that the introduction of experimentation does not change the results of the present paper, at least not for two person games. If a two person game has multiple minimal curb sets, experimentation will not yield the selection of a particular one: the limiting distribution puts positive weight on all states that are ergodic under the unperturbed process. The intuition behind this result is that only one mistake by one player is necessary in order to move the system from one ergodic set to another. When the game has more than two players, it might happen that a particular minimal curb set is selected. One can characterize the selected minimal curb set graph-theoretically.

In order to prove these results formally, we would have to recall the essential definitions and theorem from Young (1993). We refer the reader to the original paper for a formal treatment. We will just illustrate the result by means of a few examples.

Consider again the following game.

	L	R
T	2,2	0,0
B	0,0	1,1

Figure 10.

As we have seen before this game has two curb retracts, $\{TL\}$ and $\{BR\}$, and the Markov chain representing the learning process has two ergodic sets, $\{(TL, \ldots, TL)\}$ and $\{(BR, \ldots, BR)\}$. Suppose we are in state $h^{TL} = (TL, \ldots, TL)$ and player 1 makes a mistake (with probability ϵ) and plays B, so that the system moves up to state $(TL, \ldots, TL, BL)$. From the latter state the system can move, without making any further mistakes, to $(TL, \ldots, TL, BL, TR)$, to $(TL, \ldots, TL, BL, TR, BR)$, and finally to $(BR, \ldots, BR) = h^{BR}$. Hence, only one mistake is needed to move the system from h^{TL} to h^{BR}. Similarly, only one mistake is needed to move the system from h^{BR} to h^{TL}. Since the mistake probabilities are of the same order, the limiting distribution puts positive weight on both ergodic states.

It is not difficult to see that for two person games the experimentation can never select one particular curb retract: it is always the case that only one mistake is needed to move the system from one ergodic set to another one.

This result is in contrast with Young (1993). In Young (1993) the players also have information about play in the recent history: Every player draws a sample of m plays out of the plays of the most recent K periods, without replacement. Then players play a best reply in a fictitious play fashion. That is, they play a best response against the strategy combination that corresponds with the empirical frequency of strategies in the sample they draw. Let us illustrate this difference with the game from Figure 10. If player 1 makes a mistake the system can move from h^{TL} to state $(TL, \ldots, TL, BL)$. But if then nobody makes a mistake anymore, the system will move back to h^{TL}, if the sample size is at least 2: in every sample there will be at least as many T's as there will be B's. Hence, player 2 will always play L in the next period (unless he makes a mistake).

Now let us consider the following three person game, where each player i chooses between a_i and b_i.

	a_2	b_2
a_1	1,1,1	0,0,0
b_1	0,0,0	1,1,1

a_3

	a_2	b_2
a_1	0,0,0	1,1,1
b_1	1,1,1	2,2,2

b_3

Figure 11.

This game has two curb retracts, namely $\{a_1a_2a_3\}$ and $\{b_1b_2b_3\}$. Consider the case with $K = 2$. The ergodic sets are $A = \{(a_1a_2a_3, a_1a_2a_3)\}$ and $B = \{(b_1b_2b_3, b_1b_2b_3)\}$.

To move the system from A to B only one mistake is necessary. For example, the system could evolve as follows:[8]

$$\begin{array}{ccccc} (a_1a_2a_3, a_1a_2a_3) & \xrightarrow{1} & (a_1a_2a_3, a_1a_2b_3) & \xrightarrow{0} & (a_1a_2b_3, b_1b_2a_3) \\ & & & & \downarrow 0 \\ & & (b_1b_2b_3, b_1b_2b_3) & \xleftarrow{0} & (b_1b_2a_3, b_1b_2b_3) \end{array}$$

To move the system from B to A at least two mistakes are necessary. For example, the system could evolve as follows:

$$\begin{array}{ccccc} (b_1b_2b_3, b_1b_2b_3) & \xrightarrow{2} & (b_1b_2b_3, a_1a_2b_3) & \xrightarrow{0} & (a_1a_2b_3, b_1b_2a_3) \\ & & & & \downarrow 0 \\ & & (a_1a_2a_3, a_1a_2a_3) & \xleftarrow{0} & (b_1b_2a_3, a_1a_2a_3) \end{array}$$

Theorem 4 of Young (1993) then implies that the limiting distribution will put all weight on (b, b).

8 Concluding remarks

We have considered learning processes where the players have a bounded memory and play best replies against past play. The importance of the bounded memory can be elucidated by comparing our learning process with Milgrom and Roberts (1991). In general they consider games with compact strategy sets that are played continuously. Translated to the context of a two player finite normal form game which is played repeatedly at discrete points in time, they define a sequence of plays $\{s(t)\}_{t=0}^{\infty}$ to be

[8]The number above the arrows denotes the number of mistakes involved.

consistent with adaptive learning if for all $\hat{t}$, there exists a $\bar{t}$ such that for all $t \geq \bar{t}$, $s(t+1) \in \mathrm{BR}(\mathcal{B}^{ind}(\{s(\hat{t}), s(\hat{t}+1), \ldots, s(t)\}))$. We could similarly define this sequence to be *consistent with learning with bounded memory*, if there exists $K \in \mathbb{N}$ such that for all t, $s(t+K) \in \mathrm{BR}(\mathcal{B}^{ind}(\{s(t), s(t+1), \ldots, s(t+K-1)\}))$. This definition illustrates the similarity between the present paper and Milgrom and Roberts (1991).

Consider for example the pure coordination game of Figure 1. The sequence $TR, BL, TR, BL, TR, \ldots$ satisfies both definitions of consistency. However, the finiteness of the memory and of the strategy space allows us to obtain a finite Markov chain, from which we can compute that the probability of obtaining the above sequence is zero: Only sequences with tails $TL, TL, TL, \ldots$ or $BR, BR, BR, \ldots$ are obtained with positive probability.

Milgrom and Roberts (1991) show that sequences that are consistent with adaptive learning will eventually lie within the set of serially undominated strategies, which is a superset of the set of rationalizable strategies. They give some examples of games with strategic complementarities where this set is a singleton, which implies that these sequences must converge to the unique equilibrium. We get the same results in these games because the set of curb strategies is a subset of the set of rationalizable strategies. But we get similar results in some games where the set of rationalizable strategies is big. In every game that has a unique and strict equilibrium $\bar{s}$, $\{\bar{s}\}$ is the unique minimal curb set. Hence, in such games our learning process leads the players (with probability 1) to the unique equilibrium. An example of such a game is given in Figure 6, where all strategies are rationalizable.

Another example is the discretized version of the following three player Cournot oligopoly game. Player i chooses to produce q_i at zero costs to maximize $q_i(A - q_1 - q_2 - q_3)$. The unique (and strict) equilibrium is $(A/4, A/4, A/4)$. The set of rationalizable strategies is $[0, A/2] \times [0, A/2] \times [0, A/2]$.

References

Balkenborg, D. (1992). "The properties of persistent retracts and related concepts", Ph.D. thesis (University of Bonn).

Balkenborg, D. (1993). "Strictness, evolutionary stability and repeated games with common interests", CARESS working paper 93-20.

Basu, K. and J.W. Weibull (1991). "Strategy subsets closed under rational behavior", *Economics Letters* **36**, 141-146.

Bernheim, B.D. (1984). "Rationalizable strategic behavior", *Econometrica* **52**, 1007-1028.

Blume, A. (1993a). "Neighborhood stability in sender-receiver games", mimeo.

Blume, A. (1993b). "Communication, risk and efficiency in games", mimeo.

Brown, G.W. (1951). "Iterative solution of games by fictitious play", *Activity analysis of production and allocation*. New York: Riley.

van Damme, E. and S. Hurkens (1993). "Commitment robust equilibria and endogenous timing", CentER discussion paper no 9356.

Fudenberg, D. and D. Kreps (1988). "A theory of learning, experimentation, and equilibrium in games", mimeo.

Harsanyi, J. and R. Selten (1988). "A general theory of equilibrium selection in games". Cambridge, Mass.: MIT Press.

Hurkens, S. (1993). "Multi-sided pre-play communication by burning money", CentER discussion paper no 9319.

Kalai, E. and D. Samet (1984). "Persistent equilibria in strategic games", *International Journal of Game Theory* **13**, 129-144.

Kalai, E. and D. Samet (1985). 'Unanimity games and Pareto optimality", *International Journal of Game Theory* **14**, 41-50.

Kandori, M., G.J. Mailath and R. Rob (1993). "Learning, mutation, and long run equilibria in games". *Econometrica* **61**, 29-56.

Kemeny, J. and J. Snell (1976). "Finite Markov chains", 1960. Springer-Verlag, New York Heidelberg Berlin.

Milgrom, P. and J. Roberts, (1991). "Adaptive and sophisticated learning in normal form games", *Games and Economic Behavior* **3**, 82-100.

Pearce, D.G. (1984). "Rationalizable strategic behavior and the problem of perfection", *Econometrica* **52**, 1029-1050.

Robinson, J. (1951). "An iterative method of solving a game", *Ann. Math.*, **54**, 296-301.

Samuelson, L. (1994). "Stochastic stability in games with alternative best replies", forthcoming, *Journal of Economic Theory.*

Young, H.P. (1993). "The evolution of conventions", *Econometrica* **61**, 57-84.

Discussion Paper Series, CentER, Tilburg University, The Netherlands:

(For previous papers please consult previous discussion papers.)

No.	Author(s)	Title
9330	H. Huizinga	The Welfare Effects of Individual Retirement Accounts
9331	H. Huizinga	Time Preference and International Tax Competition
9332	V. Feltkamp, A. Koster, A. van den Nouweland, P. Borm and S. Tijs	Linear Production with Transport of Products, Resources and Technology
9333	B. Lauterbach and U. Ben-Zion	Panic Behavior and the Performance of Circuit Breakers: Empirical Evidence
9334	B. Melenberg and A. van Soest	Semi-parametric Estimation of the Sample Selection Model
9335	A.L. Bovenberg and F. van der Ploeg	Green Policies and Public Finance in a Small Open Economy
9336	E. Schaling	On the Economic Independence of the Central Bank and the Persistence of Inflation
9337	G.-J. Otten	Characterizations of a Game Theoretical Cost Allocation Method
9338	M. Gradstein	Provision of Public Goods With Incomplete Information: Decentralization vs. Central Planning
9339	W. Güth and H. Kliemt	Competition or Co-operation
9340	T.C. To	Export Subsidies and Oligopoly with Switching Costs
9341	A. Demirgüç-Kunt and H. Huizinga	Barriers to Portfolio Investments in Emerging Stock Markets
9342	G.J. Almekinders	Theories on the Scope for Foreign Exchange Market Intervention
9343	E.R. van Dam and W.H. Haemers	Eigenvalues and the Diameter of Graphs
9344	H. Carlsson and S. Dasgupta	Noise-Proof Equilibria in Signaling Games
9345	F. van der Ploeg and A.L. Bovenberg	Environmental Policy, Public Goods and the Marginal Cost of Public Funds
9346	J.P.C. Blanc and R.D. van der Mei	The Power-series Algorithm Applied to Polling Systems with a Dormant Server
9347	J.P.C. Blanc	Performance Analysis and Optimization with the Power-series Algorithm

No.	Author(s)	Title
9348	R.M.W.J. Beetsma and F. van der Ploeg	Intramarginal Interventions, Bands and the Pattern of EMS Exchange Rate Distributions
9349	A. Simonovits	Intercohort Heterogeneity and Optimal Social Insurance Systems
9350	R.C. Douven and J.C. Engwerda	Is There Room for Convergence in the E.C.?
9351	F. Vella and M. Verbeek	Estimating and Interpreting Models with Endogenous Treatment Effects: The Relationship Between Competing Estimators of the Union Impact on Wages
9352	C. Meghir and G. Weber	Intertemporal Non-separability or Borrowing Restrictions? A Disaggregate Analysis Using the US CEX Panel
9353	V. Feltkamp	Alternative Axiomatic Characterizations of the Shapley and Banzhaf Values
9354	R.J. de Groof and M.A. van Tuijl	Aspects of Goods Market Integration. A Two-Country-Two -Sector Analysis
9355	Z. Yang	A Simplicial Algorithm for Computing Robust Stationary Points of a Continuous Function on the Unit Simplex
9356	E. van Damme and S. Hurkens	Commitment Robust Equilibria and Endogenous Timing
9357	W. Güth and B. Peleg	On Ring Formation In Auctions
9358	V. Bhaskar	Neutral Stability In Asymmetric Evolutionary Games
9359	F. Vella and M. Verbeek	Estimating and Testing Simultaneous Equation Panel Data Models with Censored Endogenous Variables
9360	W.B. van den Hout and J.P.C. Blanc	The Power-Series Algorithm Extended to the *BMAP/PH/*1 Queue
9361	R. Heuts and J. de Klein	An (s,q) Inventory Model with Stochastic and Interrelated Lead Times
9362	K.-E. Wärneryd	A Closer Look at Economic Psychology
9363	P.J.-J. Herings	On the Connectedness of the Set of Constrained Equilibria
9364	P.J.-J. Herings	A Note on "Macroeconomic Policy in a Two-Party System as a Repeated Game"
9365	F. van der Ploeg and A. L. Bovenberg	Direct Crowding Out, Optimal Taxation and Pollution Abatement
9366	M. Pradhan	Sector Participation in Labour Supply Models: Preferences or Rationing?

No.	Author(s)	Title
9367	H.G. Bloemen and A. Kapteyn	The Estimation of Utility Consistent Labor Supply Models by Means of Simulated Scores
9368	M.R. Baye, D. Kovenock and C.G. de Vries	The Solution to the Tullock Rent-Seeking Game When $R > 2$: Mixed-Strategy Equilibria and Mean Dissipation Rates
9369	T. van de Klundert and S. Smulders	The Welfare Consequences of Different Regimes of Oligopolistic Competition in a Growing Economy with Firm-Specific Knowledge
9370	G. van der Laan and D. Talman	Intersection Theorems on the Simplotope
9371	S. Muto	Alternating-Move Preplays and vN - M Stable Sets in Two Person Strategic Form Games
9372	S. Muto	Voters' Power in Indirect Voting Systems with Political Parties: the Square Root Effect
9373	S. Smulders and R. Gradus	Pollution Abatement and Long-term Growth
9374	C. Fernandez, J. Osiewalski and M.F.J. Steel	Marginal Equivalence in v-Spherical Models
9375	E. van Damme	Evolutionary Game Theory
9376	P.M. Kort	Pollution Control and the Dynamics of the Firm: the Effects of Market Based Instruments on Optimal Firm Investments
9377	A. L. Bovenberg and F. van der Ploeg	Optimal Taxation, Public Goods and Environmental Policy with Involuntary Unemployment
9378	F. Thuijsman, B. Peleg, M. Amitai & A. Shmida	Automata, Matching and Foraging Behavior of Bees
9379	A. Lejour and H. Verbon	Capital Mobility and Social Insurance in an Integrated Market
9380	C. Fernandez, J. Osiewalski and M. Steel	The Continuous Multivariate Location-Scale Model Revisited: A Tale of Robustness
9381	F. de Jong	Specification, Solution and Estimation of a Discrete Time Target Zone Model of EMS Exchange Rates
9401	J.P.C. Kleijnen and R.Y. Rubinstein	Monte Carlo Sampling and Variance Reduction Techniques
9402	F.C. Drost and B.J.M. Werker	Closing the Garch Gap: Continuous Time Garch Modeling
9403	A. Kapteyn	The Measurement of Household Cost Functions: Revealed Preference Versus Subjective Measures

No.	Author(s)	Title
9404	H.G. Bloemen	Job Search, Search Intensity and Labour Market Transitions: An Empirical Exercise
9405	P.W.J. De Bijl	Moral Hazard and Noisy Information Disclosure
9406	A. De Waegenaere	Redistribution of Risk Through Incomplete Markets with Trading Constraints
9407	A. van den Nouweland, P. Borm, W. van Golstein Brouwers, R. Groot Bruinderink, and S. Tijs	A Game Theoretic Approach to Problems in Telecommunication
9408	A.L. Bovenberg and F. van der Ploeg	Consequences of Environmental Tax Reform for Involuntary Unemployment and Welfare
9409	P. Smit	Arnoldi Type Methods for Eigenvalue Calculation: Theory and Experiments
9410	J. Eichberger and D. Kelsey	Non-additive Beliefs and Game Theory
9411	N. Dagan, R. Serrano and O. Volij	A Non-cooperative View of Consistent Bankruptcy Rules
9412	H. Bester and E. Petrakis	Coupons and Oligopolistic Price Discrimination
9413	G. Koop, J. Osiewalski and M.F.J. Steel	Bayesian Efficiency Analysis with a Flexible Form: The AIM Cost Function
9414	C. Kilby	World Bank-Borrower Relations and Project Supervision
9415	H. Bester	A Bargaining Model of Financial Intermediation
9416	J.J.G. Lemmen and S.C.W. Eijffinger	The Price Approach to Financial Integration: Decomposing European Money Market Interest Rate Differentials
9417	J. de la Horra and C. Fernandez	Sensitivity to Prior Independence via Farlie-Gumbel-Morgenstern Model
9418	D. Talman and Z. Yang	A Simplicial Algorithm for Computing Proper Nash Equilibria of Finite Games
9419	H.J. Bierens	Nonparametric Cointegration Tests
9420	G. van der Laan, D. Talman and Z. Yang	Intersection Theorems on Polytopes
9421	R. van den Brink and R.P. Gilles	Ranking the Nodes in Directed and Weighted Directed Graphs
9422	A. van Soest	Youth Minimum Wage Rates: The Dutch Experience

No.	Author(s)	Title
9423	N. Dagan and O. Volij	Bilateral Comparisons and Consistent Fair Division Rules in the Context of Bankruptcy Problems
9424	R. van den Brink and P. Borm	Digraph Competitions and Cooperative Games
9425	P.H.M. Ruys and R.P. Gilles	The Interdependence between Production and Allocation
9426	T. Callan and A. van Soest	Family Labour Supply and Taxes in Ireland
9427	R.M.W.J. Beetsma and F. van der Ploeg	Macroeconomic Stabilisation and Intervention Policy under an Exchange Rate Band
9428	J.P.C. Kleijnen and W. van Groenendaal	Two-stage versus Sequential Sample-size Determination in Regression Analysis of Simulation Experiments
9429	M. Pradhan and A. van Soest	Household Labour Supply in Urban Areas of a Developing Country
9430	P.J.J. Herings	Endogenously Determined Price Rigidities
9431	H.A. Keuzenkamp and J.R. Magnus	On Tests and Significance in Econometrics
9432	C. Dang, D. Talman and Z. Wang	A Homotopy Approach to the Computation of Economic Equilibria on the Unit Simplex
9433	R. van den Brink	An Axiomatization of the Disjunctive Permission Value for Games with a Permission Structure
9434	C. Veld	Warrant Pricing: A Review of Empirical Research
9435	V. Feltkamp, S. Tijs and S. Muto	Bird's Tree Allocations Revisited
9436	G.-J. Otten, P. Borm, B. Peleg and S. Tijs	The MC-value for Monotonic NTU-Games
9437	S. Hurkens	Learning by Forgetful Players: From Primitive Formations to Persistent Retracts

P.O. BOX 90153, 5000 LE TILBURG, THE NETHERLANDS